Berkeley
Economic
Review

Berkeley Economic Review

UC Berkeley's Premier Undergraduate Economics Journal

Volume XIII

Content

Research Papers

Cover

Amy Ai

Policies

Submission Policy

Format: Please format your submission as a Microsoft word document or Latex file. We would prefer submissions double-spaced, 12-point, Times New Roman font. Please use the Chicago Manual of Style. Please name your file using the following format: "FirstnameLastname-PaperTitle".
Length: No maximum length for theses or original research papers.
Plagiarism: We maintain a strict zero tolerance policy on plagiarism. According to the University of North Carolina, plagiarism is "the deliberate or reckless representation of another's words, thoughts, or ideas as one's own without attribution in connection with submission of academic work, whether graded or otherwise." For more information about plagiarism-what it is, its consequences, and how to avoid it— please see the UNC's website.
Citations: Use Chicago format for in-text citations and your works cited page. Include a separate works cited page at the end of your work.
Send to: editor@econreview.berkeley.edu with the subject line "Fall 2024 Journal Submission: [Name], [Paper Title]"

Please direct any questions you may have to the same email address as above. We will reach out after the above deadlines regarding your acceptance status to our journal. Please note that we no longer accept any op-eds, essays, or articles for consideration.

Disclaimers

The views published in this journal are those of the individual authors or speakers and do not necessarily reflect the position or policy of Berkeley Economic Review staff, the UC Berkeley Economics Department and faculty, or the University of California, Berkeley in general.

Copyright Policy

Staff

Editors in Chief

Larry Lin
Pallavi Murthy

Research and Editorial

Department Heads
Gia De Mayo
Ramya Sridhar

Writers & Editors
Dima Zayaruzny
Angela Chen
Aayush Singh
Saniya Pendharkar
Kiet Hoa
Lara Ramirez
Tucker Gauss
Griffin Shufeldt
Alkis Toutziaridis
Esma Ozgun
Nelson Law
Emma-Jane Burns
Arda Tuncturk
Anne Wu
Selina Yang
Nelson
Cole Yap
Felix Zhang
Jack Yang
Aidan Morgan Chan
Isabel Chen
Sharla Sabaruddin
Sophie Niu
Davina Yashar
Andrew Ho
Keshav Kumar
Shashi Balasubramanian
Emily Vu
Miriam Klaczynska
Ari Gomes
Rebecca Peng

Head of Professional Development

Syrak Micael

Staff

Peer Review

Department Head
Adam Staley

Associates
Pietro Risso
Matthew Liu
Greg Simitian
Elise Yi Gao
Melvin Adams
Nicolas Sarullo
Sarah Siemsgluess
Austin Zhu
Elise Yi Gao
Stella Jia
Nick
Neha Muthiah
Eva Heyl
Thomas Hung
Murari Ganesan
Aryan Shafat
Michael Athanassiadis

Head of Social

Mia Depena

Head of Finance

Markus Leung

Layout and Design

Department Head
Dylan Margolis

Associates
Amy Ai
Citlali Gonzalez
Giulia Pappalardo
Cynthia Shi

Marketing

Head
Amy Hoefs

Associates
Victoria Lee
Brandon Vecchio
John Belichak
Elizabeth Nichols
Priscilla Jun

Staff

Outreach

Department Head
Melvin Adams

Associates
Ella Marsh
Maneesh Vallurupalli
Vismaad Randhawa
Arnav Machavarapu
Vittal Vasudevan
Shining Lu

Empirical Analysis

Department Heads
Griffin Shufeldt
Nicolas Sarullo

Associates
Vidya Venkatachalam
Anishka Banerjee
Anushka Mangal
Naomi Kung
Sarah Dong
Ryan Raghavan
Aneesh Kudrimoti
Ria Bhandarkar

Editor's Note

Dear BER Reader,

Since its inception in 2016, the Berkeley Economic Review has been a beacon for showcasing top undergraduate economics research. Our journey began with a vision to publish innovative and impactful contributions while nurturing the talents of future economists. Today, BER thrives with over 90 dedicated members, hundreds of submissions, and tens of thousands of readers from around the globe.

In a year marked by significant events such as the ongoing Russia-Ukraine war, the revolutionary launch of GPT-4, the impactful Hollywood strike, and the tragic Titan submersible implosion, we are reminded of the ever-changing landscape in which we operate. Amid these dynamic shifts, the unwavering dedication of our authors and team has been our cornerstone. Their resilience and creativity shine through in times of geopolitical challenges, economic uncertainties, and societal transformations.

Our Peer Review team has meticulously selected exceptional undergraduate research that addresses some of the most pressing issues of our time. This volume includes insightful analyses on the causal relationship between inequality within housing prices and crime rates in London and Wales, the effects of increased policing on crime, and the legislative impacts on the oil industry. We also explore voter behavior across partisan lines, the influence of mobile phone access for women on household consumption in India, and the rise in school enrollments following the abolition of the army in Costa Rica.

With great enthusiasm, we are thrilled to unveil the 13th volume of the Berkeley Economic Review, a testament to the enduring spirit of scholarly inquiry and the pursuit of excellence.

Sincerely,

Pallavi Murthy & Larry Lin
Editors-in-Chief
Berkeley Economic Review

Estimating the Effect of Local Income and Wealth Inequality on Crime Rates in England and Wales

May 15, 2022

Abstract

This paper explores the relationship between localised inequality and crime in England and Wales from 2014-2021. Using logistic regression and comparing the effects observed over the Middle-layer Super Output Area and Local Authority scales, a statistically significant correlation was found on the Local Authority scale. However, first-differencing the models to establish causality removes this effect. This suggests that the correlation is caused by fixed-effects outside of the model and not small area income inequality. When breaking crimes down into property, violent, and other crime, an effect was only found in other crime. Investigating the effect of income relative to the mean of one's neighbours on a middle-super output area level gives a negative effect within cross sectional analysis, and time-series analysis shows that a negative effect in other crime is hiding a positive effect in property crime. However this effect is very small, and not economically significant. The novel use of house prices to measure inequality gives far stronger results, finding a statistically significant relationship between house price inequality and all types of crime, when first-differenced, suggesting a causal relationship between the two variables.

Content

1 Introduction

Predictive models of crime would offer police forces a tool to ensure the most efficient expenditure on crime deterrence [25]. The ability to see where crimes are most likely to take place before they occur would allow police forces to patrol these areas more heavily, using their limited labour and capital resources with maximum preventative effect. While these models have been described as problematic in the past, they offer such scope for benefits that we must strive to find a model which prevails, overcoming their issues, the key to which is using the correct input data, ideally outside of reports of recent crimes [7, 21]. Thus, this paper will focus on the 'demand-side' of crime, or the rates of victim hood, since deterrent behaviour comes from the protection of likely targets and not the finding of likely criminals [11, 25].

Inequality's relationship with crime is a contested one. In theory, local inequality leads to increased closeness of agents with high incomes, who are more profitable targets for crime, and agents with low incomes, with a low opportunity cost of committing crime [11, 29]. This closeness leads to increased contact, creating more opportunities for criminal behaviour. When this relationship is researched on a national scale, the results are clear; higher levels of inequality cause higher levels of crime [20, 38]. However, when this relationship is explored within smaller areas, more compatible with normal criminal behaviour, this relationship is far less obvious, many papers finding no significant relationship, some with conflicted results, and some finding a statistically significant positive relationship [11].

This paper aims to resolve the conflict between the theoretical and empirical evidence regarding a causal relationship between local inequality and crime within the UK. England and Wales are a great option for study as the crime data is very open and fine-grained. This means that exploring the relationship on a local level is far more achievable than in other nations. A small area explanation of the supposedly causal national relationship has stronger backing for economic motivations behind crime. This comes from the fact that as the proximity of high value victims and low cost criminals increases, the opportunity for criminal profits increase. Criminals have been shown to commit crime within an average of 2 miles from their place of residence, meaning that over a national scale the presence of heterogeneous inequality would be expected to dampen the effects of national inequality [11].

The papers which have attempted to answer this question in the past in the UK have been on scales smaller than that which this paper will explore. Whitworth [43] performed research on the same sizes of areas in the UK, middle-layer super output areas (MSOAs), exploring the relationship between the two variables, with measures of inequality formed from an aggregation of these areas. However, Whitworth [43] only researched these effects within London and South Yorkshire, whereas this research will cover the entirety of England and Wales. Another innovation of this current paper is that we capture the visible differences in incomes in small areas by observing the variation of house prices within an area instead of raw income. Visible income differences are a more valuable metric as those looking to commit crime will not know the income levels of those around them, only able to form a prior belief of the value of the crime based on the visible wealth of the victim.

2 Literature Review

2.1 Inequality and Crime

This theory has been found to explain international heterogeneity in crime by Fajnzylber et al. [20], who discovered a positive relationship between inequality and homicide and robbery when looking at panel data. However, Stack [39] found a negative effect between the Gini index of countries and the crime rate, but this is explained by his use of cross sectional data. This means that poorer nations with both lower rates of inequality and greater rates of under-reporting to lead to an incorrect conclusion. This sort of error is present in many papers on this topic [38].

Multiple papers have used differences in inequality to explain national variation in crime rates. Doyle et al. [17], Saridakis [35], and Chintrakarn and Herzer [9] compared crime rates across US states using panel data. Their results were weak, with the only statistically significant result being Saridakis [35] finding a positive relationship between inequality and murder. Chintrakarn and Herzer [9] explained the lack of connection by theorising that a variable was missing from the equation; regions with more inequality had greater consumption of personal crime prevention products, examples of which may be burglar alarms and bodyguards. This would mean that a lower success rate would offset the heightened propensity to commit crime. Corvalan and Pazzona [12] explored the effects in South Africa by comparing police municipalities, finding no relationship between the two variables, before also explaining their lack of correlation on the unobserved consumption of crime protection.

Another area of the world where research into this relationship was undertaken was in the USA. This was more popular early on in research into crime rates as it ensured relatively consistent recording methods across a large population. Ehrlich [19] completed his paper, having introduced the theoretical relationship between inequality and crime, with an empirical exploration of the relationship finding a positive and significant link.

Many researchers have examined the relationship between inequality and crime rates on the scale of American Standard Metropolitan Statistical Areas (SMSAs), encompassing entire cities, with mixed results. Eberts [18] compared total crime across SMSAs, finding a positive relationship between crime and inequality. Danziger and Wheeler [13] established a positive correlation between the Gini index and burglary, robbery, and assault. However, Blau and Blau [4] found no significant effect of the Gini index on any crime other than murder and rape. Williams [44] found no relationship between murder and the Gini index.

Researchers have recently investigated the relationship within regions smaller than whole cities. These smaller areas are more compatible with the economic incentives surrounding income inequality as criminals tend not to travel far to commit crimes, averaging 2 miles from their home [11]. Kelly [26] found positive effects within US counties. Stucky et al. [40] explored the relationship of the inequality of neighbourhoods, finding positive relationships between the Gini coefficient and crime, present in all types of crime apart from homicide. Brush [6] conducted a study on US counties, uncovering a positive cross-sectional correlation. However this appears spurious since when the data is first-differenced, he gets a negative result.

Since income inequality data is often limited for small areas, some researchers have used spatial models, comparing the income levels of adjacent neighbourhoods to give a relative income of a target area compared to the mean of its neighbours. Metz and Burdina [28], after finding no significant relationship between the Gini coefficient and the crime rate in census block groups, used this method and established that block groups that were poorer than their neighbours experienced lower crime rates, but with no significant effect from being more affluent than one's neighbours, probably due to increased security. When they explored the impact of the size of the difference from a block's poorest or richest neighbour, they unearthed an effect on the other end, finding that the only

significant result was that the greater the difference in income from a block's most impoverished neighbour, the greater the level of crime that is experienced. Scorzafave and Soares [36] investigated the relationship in Sao Paulo, finding an elasticity of property crime with the Gini index to be 1.46.

The most similar piece of research that we have come across is Whitworth [43], who studied the effects of inequality on the crime rate in MSOAs in London and South Yorkshire, finding a significant relationship between the two variables, with the effect being greater in London than in South Yorkshire. This is what one might expect given that the MSOAs are much smaller in area in London, leading to a greater level of proximity between agents. The approach taken is similar to that in Metz and Burdina [28], comparing the regions to their neighbours rather than measuring inequality within that region. This leads to a large amount of crossover in the inequality measure. One region with an extreme income level would lead to all the areas around it being recorded with terribly similar inequality rates. The measurement also has little respect for the size of the neighbouring area, large regions of rich countryside in South Yorkshire affect the edges of cities. These rich areas leave the city edges as very unequal when their more densely packed neighbour closer to city centre will be a greater influence on the mindsets of inhabitants. There is also no accounting done for the distance between the areas, only if they are neighbours, or neighbours' neighbours etc. This means that this spatial model actually doesn't measure space, instead measuring adjacency. This matters in London, where the MSOAs are more densely packed since their boundaries are set for relatively consistent population sizes, and likely to effect MSOAs that may not be adjacent but are still close.

2.2 Other Causes of Crime

The two most important factors found to influence the crime rate are income and poverty. These are the variables most obviously related to the drive to commit crime from the economic theory first presented in Becker [3]. Lower incomes will lower opportunity costs for those choosing to commit crime rather than earn through legal employment, increasing u. There is also a reduction in the loss of earnings from imprisonment, reducing f. While the two of these are hard to test together due to the natural collinearity problems, poverty is the better predictor of crime rates [24].

Another variable that has been determined to effect on the rate of crime is the popula tion density of an area. This affects the probability of prosecution as it is more likely that there are others around to witness the crime taking place. However, there is an opposing effect of high population density, where, as the density increases, residents know each other less. This makes criminals who have been seen committing a crime less identifiable, but within social disorganisation theory, it also reduces the control that society has on the behaviour of its members [37, 34]. If individuals know each other less well, they will feel less guilt committing crime against the citizens around them. This means that while p is increasing in population density, u is increasing as the 'guilt cost' of crime decreases.

Young people are more likely to be involved in crime, both as victims and as perpetrators [10]. Therefore, an area with a younger population would be expected to experience a higher crime rate. This means that the proportion of young people within an area will need to be accounted for within the model.

Another variable theoretically linked with crime rates is the number of police in an area, as Becker [3] suggests that a criminal needs to estimate the chance of their illegal activity being successful to estimate its value. A handful of papers cast doubt over the effect of police spending but their conclusions were unreliable due to the endogenous nature of police spending, rising in high crime areas [45, 8]. Eventually, Levitt [27] put the issue to rest, brilliantly using the cyclical nature of police expenditure during election cycles to remove the endogeneity. This backed up the finding of Ehrlich [19] and showed that police expenditure did cause crime rates to fall. However, since this paper takes place over a short time period in a nation with only one police force, police expenditure is assumed to be consistent.

3 Data

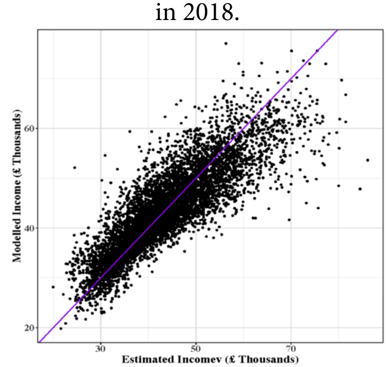
For this research, the data on crime is compiled from data.police.uk [14]. These data contain all crimes reported, within England and Wales, with the month that the crime was reported, and the street upon which the crime reportedly occurred. Reports shown to be false reports when investigated are removed. Some locations are adjusted in order to ensure that the victims are not identifiable. Due to limitations in other data, this will be aggregated across time annually. Geographic aggregation will be

completed to two levels: MSOA and Local Authority (LA). These areas were defined in the 2011 census [32][1].

My source of income data comes is Office for National Statistics (UK) [30]. This data provides the average annual income of MSOAs over the required years. This again can be simply aggregated. However, due to the single data points for each area, income inequality for MSOAs is not achievable, so we will use the spatial analysis method used by Whitworth [43] and Metz and Burdina [28], comparing the incomes of areas to those of their neighbours.

To generate another measure of inequality house pricing data, provided by Zoopla Limited © [47]. In order to find income inequality using this data, a sample of 350,000 sales listing prices and the same number of listed rental prices were taken from each year. These were then sorted into MSOAs before the Gini coefficient of the prices in that area were taken. Sorting into MSOAs after sampling gave areas an average of 90.41 price points in each year, leaving enough points to create a reliable Gini coefficient. Alongside this, since the ONS Geography [32] data finishes in 2018, overlapping with the house pricing data by only a year, in order to create a mean income value for the areas after that point, an instrumental variable was created as a replacement, created from the mean sale and rental prices in that area. This captures 71.5% of the variation in house prices as shown in Figure 1. The modelled incomes are used for analyses using MSOA Gini coefficients, whereas the true incomes are used for any analyses using income relative to neighbours.

Figure 1: The modelled area mean incomes against the true area incomes in 2018.



1 Source: Office for National Statistics licensed under the Open Government Licence v.3.0

In order to measure population density, Office for National Statistics (UK) [31] was used. This provides both the population of each Lower Super Output Area (LSOA), one level of aggregation below the desired MSOA level. In order to aggregate these, the sum of the LSOA populations was divided by the sum of the LSOA areas in that MSOA, or mathematically:

$$\frac{\sum_{LSOA \in MSOA} Population_{LSOA}}{\sum_{LSOA \in MSOA} Area_{LSOA}} = PopulationDensity_{MSOA}$$

In order to aggregate up to the LA level, the same format was used but with less numerous larger sets.

For the population data, as well as the ages of inhabitants this research uses Office for National Statistics (UK) [31]. This dataset contains the populations within each LSOA, again allowing aggregation to MSOA and LA levels. Total populations are reported in the data, along with the estimated populations of each age, down to the year. In order to find the percentage of the population in each area that is between the age of 16 and 25, the following formula was used:

$$\frac{\sum_{a=16}^{25} Population_a}{Population_{area}} * 100 = Age16 - -25(\%)_{area}$$

Finally, data on the claimant count in each area, was sourced from [16]. For this, the adjusted claimant count was used, a count which adjusts for changes in legislation, meaning that only individuals who would have qualified as claimants under the legislation of 2013 are counted, allowing comparison across time. Again, to avoid collinearity, this was adjusted to the percentage of the population of an area claiming, rather than the raw count.

4 Modelling

4.1 Underlying Model

In the model used in this paper, we assume a population of individuals with a distribution of possible legal incomes y_i. Given the loss of their current income by moving into criminal employment, this means u is a function of y and is strictly decreasing in terms of y, i.e. the opportunity cost of illegal employment rises. For the sake of this paper, it is assumed

that all other factors which contribute to O in any way is captured within the variables in the analysis. In order to capture the effects of inequality we must also model the 'demand' for crime, or the victims who 'pay' the criminals for victim-hood. This is an odd market as the consumption of crime provides only disutility. We will model the market for criminal behaviour as follows:

1. Each period, individuals within the population are paired up, assigned a possible legal income for that period y_i, and assigned a role in the game, criminal or victim.

2. The victim is assigned a value x_v. This is the amount that the victim stands to lose through the interaction, and since this is a zero-sum interaction, the amount that the criminal has to gain.

3. Criminal gets a signal from the victim, offering them some information regarding the value of x_v.

4. The criminal decides whether to commit the crime against the victim.

   (a) If the criminal goes through with the crime, they enter a lottery with two possible outcomes.

      i. They are prosecuted with probability p, receiving payoff $U_c = -f$, meaning the victim receive $U_v = y_v$.

      ii. They get away with it meaning that they receive $U_c = x_v$ with probability $1 - p$, and the victim receives $U_v = y_v - x_v$.

   (b) If the criminal decides not to commit the crime, both parties receive their incomes $U_i = y_i$.

In this model, if x_v is increasing in terms of y, then those on higher incomes become higher value targets. This means a signal related to income, such as the observable value of a house, would lead to more crimes being committed against an individual, and potential criminals who are at the margin when coming across an individual with average income, will become a criminal when interacting with an above average earner. If we assume that f and p are constant across the population, then for every y_c, there is a minimum $y_{v,commit}$ beyond which the individual will always commit the crime. The same works in the other direction, every

y_v has a corresponding maximum $y_{c,commit}$, such that any individual earning below this threshold will commit a crime against the victim. In areas of higher income inequality, there will be more pairs with low y_c and high y_v leading to higher propensity to commit crime. This means that we would expect more crimes to take place in these areas, increasing O.

This paper assumes that p and f do not change over time for individuals. There are flaws to these assumptions, for example, those in old age may not be able to serve lengthy prison sentences, creating a limit to f which reduces across a lifetime. However, the levels of these effects should remain relatively constant in the 4-year time-scale the analyses within this research take place over.

4.2 Econometric Model

The model for this paper is that the rate of crime in an area is driven by the characteristics of that district. This arises from the assumption that crime is committed by criminals near to where they live so by measuring the crime rate of an area you measure, to some extent, the propensity of its inhabitants to commit crime, and to a greater extent the likelihood that its inhabitants will be victims [10, 1]. This research models measured effect within two sizes of area. Firstly, a comparison on the scale of LAs within England and Wales, then within a finer grain, MSOAs. The average incomes of MSOAs compared to their neighbours was measured. This means that the model which we measured econometrically is as follows:

$$Y_{it} = \beta X_{it} + u_t + \varepsilon_{it}$$

Within this model, Y_{it} was the level of crime in area i at time t, β is a matrix of coefficients which is estimated in the modelling. X_{it} is the (n × 1) matrix of measurements of all of the variables in the model. The variables used in the analysis are listed in Table 1 and Table 2. We also include error terms for both that year on the whole u_t, and for that area in that year ε_{it}.

Table 1: Summary statistics of the data aggregated on the LA level for the years 2014-2018.

Statistic	N	Min	Pctl(25)	Median	Pctl(75)	Max	St. Dev.
Gini	1,029	0.01	0.06	0.07	0.09	0.17	0.02
Crimes	1,029	1,311	7,163	11,376	21,619	129,474	15,193.53
Property Crimes	1,029	387	2,202	3,572	6,603	47,913	4,752.06
Violent Crimes	1,029	195	1,082	1,759	3,412	36,352	3,163.05
Other Crimes	1,029	674	3,785	6,049	11,477	59,320	7,935.20
Crimes Per 100	1,029	3.37	6.74	8.66	11.28	32.10	3.44
Mean Income	1,029	26,809	36,173.3	40,540	46,250	66,825	7,909.03
Claimants (%)	1,029	0.70	1.22	1.74	2.48	5.55	0.89
Pop. Density	1,029	353.05	1,841.76	2,933.06	4,284.09	20,796.02	2,993.06
Population	1,029	34,410	98,127	131,819	204,525	1,141,374	113,094.10
Age 16–25 (%)	1,029	8.16	10.08	10.95	12.17	26.07	2.81

Table 2: Summary statistics table when aggregated at the MSOA level for the years 2014-2018.

Statistic	N	Min	Pctl(25)	Median	Pctl(75)	Max	St. Dev.
Income Relative to Neighbours	19,756	−27,740.4	−4,239.0	−558.6	3,328.3	38,818.2	6,126.3
Crimes	19,756	62	404.8	628	972.2	20,289	754.9
Property Crime	19,756	16	118	193	308	4,816	208.4
Violent Crime	19,756	2	46	79	144	12,844	250.7
Other Crime	19,756	24	220	340	521	8,392	360.7
Crimes Per 100	19,756	1.3	5.4	8.0	11.7	241.2	7.8
Income	19,756	17,680	34,100	40,200	47,840	96,900	10,310.2
Claimants (%)	19,750	0.3	1.0	1.7	2.8	13.7	1.5
Pop. Density	19,756	8.9	1,874.4	3,575.6	5,430.6	30,523.3	3,928.8
Population	19,756	2,242	6,742	7,859	9,097	24,969	1,797.8
Age 16–25 (%)	19,756	4.8	9.7	11.0	12.6	83.3	5.7

This research also finds the Gini-coefficients withing MSOAs using house sale and rental prices from Zoopla Limited © [47]. These are aggregated by MSOA and then the Gini coefficients from sale prices and rental prices are found separately and an average of the two is used to give the Gini for each area. The summary statistics of the data generated by this aggregation are in Table 3.

Firstly, a simple check of the correlation of the variables was performed. This allowed us to establish whether there is a relationship between income inequality and crime rates as would be expected. However, this model fails to take into consideration the fixed effects of areas in the study. To remove these fixed effects, a first-differencing model was used. This results in the removal of geographic fixed effects, meaning that a result here are far more indicative of a causal relationship. This first-differenced model is written mathematically as:

$$(Y_{it} - Y_{i(t-1)}) = \gamma(X_{it} - X_{i(t-1)}) + v_t + e_{it}$$

Table 3: Summary statistics for the data aggregated at the MSOA level, using house prices to find the Gini coefficient within MSOAs for the years 2018-2021.

Statistic	N	Min	Pctl(25)	Median	Pctl(75)	Max	St. Dev.
Gini	19,665	0.00	0.15	0.19	0.23	0.51	0.06
Crimes	19,666	0	439	678	1,047	25,028	795.72
Property Crimes	19,666	0	164	257	397	4,816	248.41
Violent Crimes	19,666	0	42	74	136	17,388	285.84
Other Crimes	19,666	0	210	327	505	8,124	352.84
Crimes Per 100	19,666	0.00	5.75	8.48	12.37	323.43	7.88
Claimants (%)	19,660	0.26	1.17	1.99	3.20	15.89	1.65
Pop. Density	19,666	8.78	1,896.88	3,599.46	5,472.41	31,867.62	4,050.16
Population	19,666	4,825	6,815	7,955	9,274	27,911	1,943.61
Age 16 25 (%)	19,666	4.52	9.19	10.38	11.75	83.27	5.65

4.3 Estimation Techniques

Using this model, the parameters, βx are firstly estimated using a maximum likelihood estimation. A Poisson distribution is assumed as the distribution of crime closely re sembles this form. As with all regressions in this paper, the model was run first on the LA level using the Gini coefficient of mean incomes of MSOAs in that LA. Since this is performed on the means of area incomes, these coefficients are very low as much of the variation is removed, but it will still capture the gaps in earnings across areas in those LAs. On the MSOA level, income inequality was generated using the income relative to the neighbouring areas. For this paper, neighbouring is defined as any MSOA within 2 miles of the border of the the target MSOA. This distance was chosen as this is the average distance between residency and crime [11]. This means that the neighbouring MSOAs are within the average travelling distance to commit a crime. Since we measure the crime that takes place inside the target MSOA, not the number of crimes committed by its residents, we expect crime rates to be higher in areas which are rich relative to their neighbours, ceteris paribus. This is because in an area with a relatively high income, committing crime will have a higher rate of reward, meaning that they should be willing to take higher risks in going through with the crime. The models were also created using the inequalities within MSOAs derived from house prices. Then, in order to establish causality, the same models were

run using the first-differencing method, treating the data as panel data. This is more appropriate for causal analysis in this case as this captures the fixed effects of each area.

5 Results

5.1 Cross-Sectional Analysis

These results are far easier to interpret visually. In order to reduce the dimensionality of any visualisations, we have used partial regression plots treating income inequality as the additional variable [41]. This allows for easy interpretation of the results, effectively showing the relationship between the independent variable we are interested in, income inequality, and the changes in the dependant variable, crime, not explained by any other variable in the model. Figure 2 shows the effect found when observing the data on the LA scale. This, along with Table 4 shows that when viewing the data in a cross sectional manner, we can view a correlation, with strong statistical significance suggesting an inequality elasticity of crime of around 5.5. Table 4 also shows that the regression results for the other variables are in the same directions as we would expect, youth, population density, and claimants increasing crime rates and mean income having a negative effect. Robustness checks for heteroskedasticity were performed, and it was found to be present, so all standard errors are White's heteroskedasticity robust standard errors [42]. This is the case for all regressions where serial correlation was found to be present.

Figure 2: Partial regression plots for the data aggregated at the LA (left) and MSOA (right) level between the years of 2014–2018.

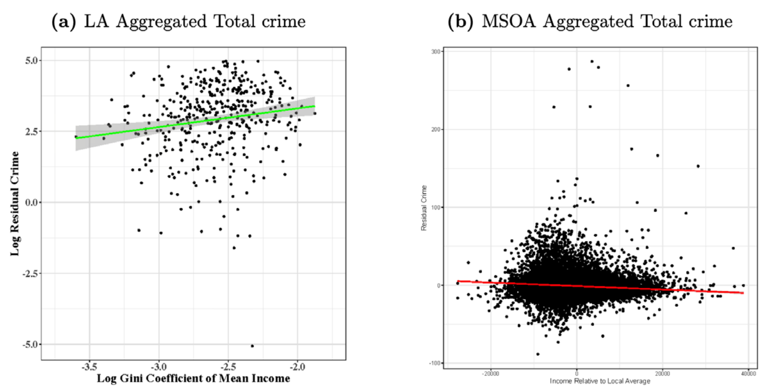


Table 4: Results of the cross-sectional regressions performed on the LA aggregated data. Since these are log-log models, the coefficients are to be interpreted as elasticities.

	Dependent variable:			
	Crimes	Property Crime	Violent Crime	Other Crime
Gini	5.546***	5.012***	5.454***	5.870***
	(0.642)	(0.697)	(0.690)	(0.662)
Claimants (%)	0.117***	0.146***	0.060	0.122***
	(0.025)	(0.027)	(0.047)	(0.024)
Pop. Density	0.0001***	0.00004***	0.0001***	0.0001***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Pop	0.00000***	0.00000***	0.00000***	0.00000***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Age 16–25 (%)	0.022***	0.015***	0.029***	0.024***
	(0.004)	(0.004)	(0.006)	(0.004)
Year 2016	0.171***	0.421***	0.068	0.082**
	(0.040)	(0.041)	(0.055)	(0.040)
Year 2018	0.313***	0.752***	0.209***	0.099**
	(0.043)	(0.046)	(0.053)	(0.043)
Mean Income	−0.00001***	−0.00001***	−0.00000	−0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Constant	8.315***	7.109***	6.144***	7.768***
	(0.144)	(0.147)	(0.249)	(0.140)
Observations	1,029	1,029	1,029	1,029
Log Likelihood	−845,245.400	−295,774.800	−185,261.000	−496,345.300
Akaike Inf. Crit.	1,690,509.000	591,567.600	370,539.900	992,708.700
Note:	White [42]'s standard errors used; *p<0.1; **p<0.05; ***p<0.01			

In the case of relative MSOA incomes, since we are viewing the rate of crime against the income of an area relative to its neighbours, in the model we would expect the high income areas to be higher value targets of crime, and the lower income areas to have greater populations of those willing to commit crime. In the model this should lead to the poorer populations in low income areas travelling to the higher income areas to commit crime, so we would expect the areas with incomes higher than their neighbours to suffer higher rates of crime. However, Figure 2 shows that this is the opposite of what we see. Here, the crime rates not predicted by the other variables are higher in the relatively low income areas than in the high income areas. This is a surprising result, especially given that the rate of crime caused by low income should be accounted for. Table 5 shows that this is a significant result. This result opposes the findings of Metz and Burdina [28], who found that blocks poorer than their neighbours had lower crime rates rather than higher, ceteris paribus, but our result is very small indeed.

This unexpected pattern continues into the analysis performed on the later years using the house price inequality data. This in fact finds a small negative correlation between crime and the Gini coefficient of the house prices. This finds the inequality elasticity of crime to be -0.046, as shown in Table 6, very very small and unexpectedly negative. The size of this is especially clear when shown graphically as in Figure 3. This again suggests that there is little to no positive relationship between crime and localised inequality. While this result is negative, it is simply too small to be important.

Theoretically, the different motivations for different types of crime may lead to the rates of different crimes being driven by different factors. Here we pull apart profit driven property crimes, hateful, tension driven violent crimes, and other crimes, mainly consisting of the consumption of illicit drugs and speeding [33, 5] . If this were the case, we would expect a positive effect on property crime and a negative effect in another category to be hiding it. The details of precisely which crimes reported in the data.police.uk [14] dataset are in each group are in Table 7. Crimes were grouped rather than broken down individually as this created numbers which were too small, with many zeros. Table 4 shows that there remains a positive correlation in property crime on the LA level. Partial regression plots for these models are in Figure 4. Also contained in these plots are the results when aggregated at the MSOA level. There are also remains no changes in the effects seen, with the negative effects of inequality in fact increasing when broken down into groups, but still inconsequentially small. The results of this breakdown are in Table 5.

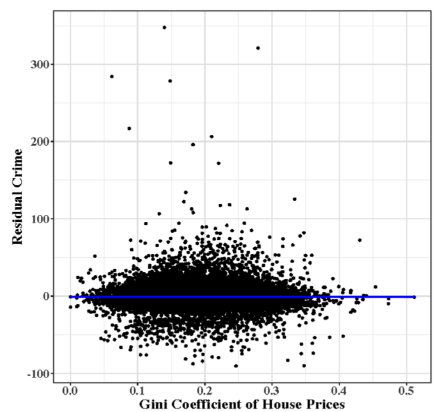
Table 5: Results of the cross-sectional regressions performed on the MSOA aggregated data. Since these are log-log models, the coefficients are to be interpreted as elasticities.

	Dependent variable:			
	Crimes	Property Crime	Violent Crime	Other Crime
Income Relative to Neighbours	−0.00001***	−0.00002***	−0.00002***	−0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Claimants (%)	0.183***	0.181***	0.235***	0.169***
	(0.006)	(0.005)	(0.011)	(0.005)
Pop. Density	0.00001***	0.00000	0.00001***	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Pop	0.0001***	0.0001***	0.0002***	0.0001***
	(0.00000)	(0.00000)	(0.00001)	(0.00000)
Age 16–25 (%)	0.015***	0.009***	0.026***	0.015***
	(0.001)	(0.001)	(0.002)	(0.001)
Year 2016	0.201***	0.442***	0.144***	0.100***
	(0.013)	(0.011)	(0.030)	(0.011)
Year 2018	0.288***	0.729***	0.203***	0.067***
	(0.013)	(0.011)	(0.031)	(0.011)
Income	0.00000*	−0.00001***	0.00002***	−0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Constant	4.622***	3.621***	1.467***	4.304***
	(0.053)	(0.041)	(0.105)	(0.046)
Observations	19,750	19,750	19,750	19,750
Log Likelihood	−1,990,036	−508,729	−1,055,088	−976,350
Akaike Inf. Crit.	3,980,090	1,017,476	2,110,193	1,952,718
Note:	White [42]'s standard errors used; *p<0.1; **p<0.05; ***p<0.01			

Table 6: The results of cross sectional model regressions using inequality numbers generated by house prices. Since there are log-log models, the coefficients are interpreted as elasticities.

	Dependent variable:			
	Crimes	Property Crime	Violent Crime	Other Crime
Gini	−0.046	−0.042	−0.548	0.284**
	(0.144)	(0.097)	(0.397)	(0.114)
Income	−0.00000	−0.00002***	0.00004***	0.00000
	(0.00000)	(0.00000)	(0.00001)	(0.00000)
Claimants (%)	0.183***	0.122***	0.317***	0.183***
	(0.012)	(0.007)	(0.035)	(0.009)
Pop. Density	0.00001***	−0.00000	0.00001*	0.00002***
	(0.00000)	(0.00000)	(0.00001)	(0.00000)
Age 16–25 (%)	0.013***	0.009***	0.022***	0.012***
	(0.001)	(0.001)	(0.002)	(0.001)
Population	0.0001***	0.0001***	0.0002***	0.0001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Year 2019	−0.028*	−0.004	0.020	−0.066***
	(0.016)	(0.010)	(0.045)	(0.013)
Year 2020	−0.380***	−0.362***	−0.757***	−0.294***
	(0.015)	(0.012)	(0.037)	(0.014)
Constant	5.141***	5.398***	1.157***	4.194***
	(0.120)	(0.067)	(0.349)	(0.083)
Observations	19,659	19,659	19,659	19,659
Log Likelihood	−2,289,187	−725,554	−1,062,392	−1,072,718
Akaike Inf. Crit.	4,578,391	1,451,127	2,124,803	2,145,454
Note:	White [42]'s standard errors used; *p<0.1; **p<0.05; ***p<0.01			

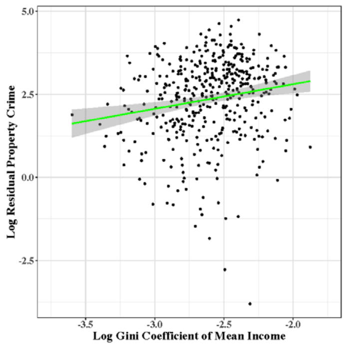
Figure 3: Partial regression plot of the correlation between income inequality and crime rates.



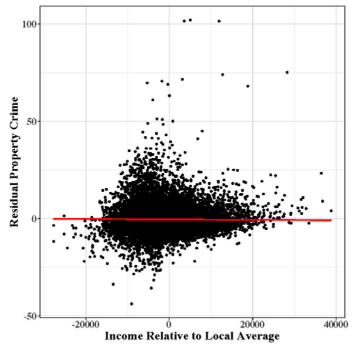
When measuring the different effects within different types of crime using the inequality of house prices, we do find there to be heterogeneous effects, as shown in Table 6 and Figure 5. Property crime and violent crime were statistically significant but these findings are not heteroskedasticity robust. High rates of violent crime will push down the prices of houses in that area, and if prices are low across an area, there will be little inequality of prices across that area. On the other hand, other crimes have a strong positive correlation, with elasticity 0.3.

Figure 4: The partial regression plots for each crime group when LA and MSOA aggregated. The left column contains the LA aggregated models and the right the MSOA aggregated models. The rows, in descending order are property crime, violent crime, and other crime.

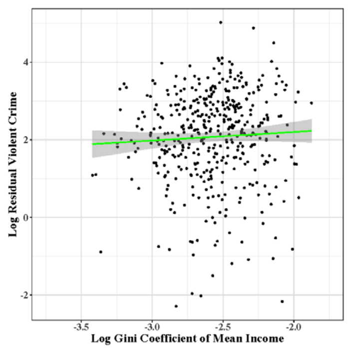
(a) LA aggregated property crime



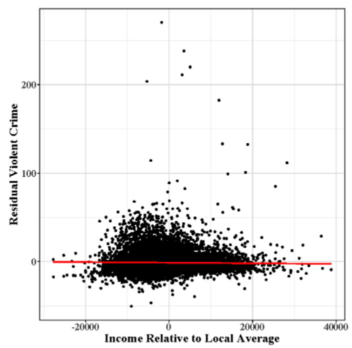
(b) MSOA aggregated property crime



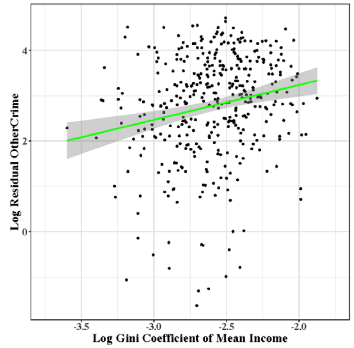
(c) LA aggregated violent crime



(d) MSOA aggregated violent crime



(e) LA aggregated other crime



20

(f) MSOA aggregated other crime

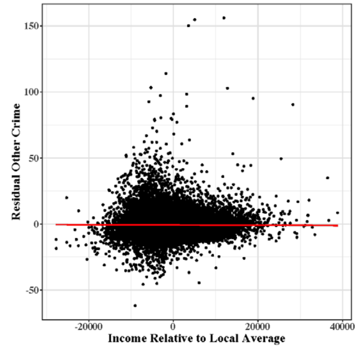
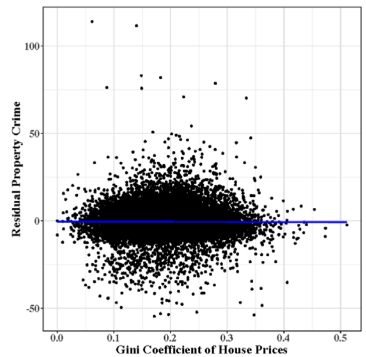
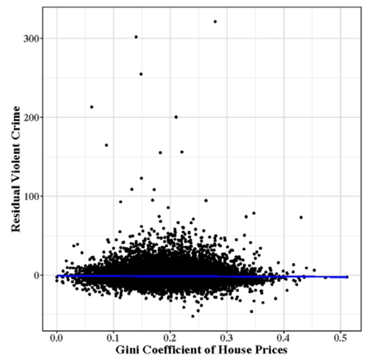


Figure 5: Plots of the partial regressions of house price inequality against residual crime, as in the table, the only significant relationship is in other crime (right), with no relationship found in property crime (left) or violent crime (middle).

(a) MSOA aggregated property crime



(b) MSOA aggregated property crime



(c) MSOA aggregated property crime

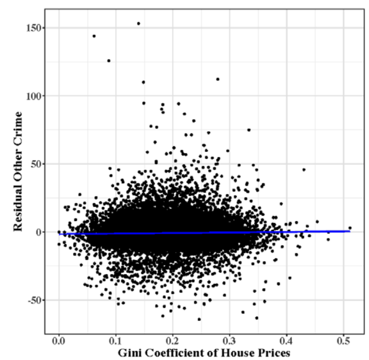


Table 7: The classifications of crimes in the dataset, and the group which they have been placed in within this research.

Crime Group	Crimes Included
Property Crime	Bicycle theft, Burglary, Robbery, Shoplifting, Theft from the person, Other theft
Violent Crime	Criminal damage and arson, Possession of weapons, Violence and sexual offences
Other Crime	Anti-Social behaviour, Drugs, Other crime, Public order, Vehicle crime

5.2 Time-Series Analysis

When we try to view the effects over time, the story is very different. Here we see that, while the effects of other variables maintain the same directional effect and remain statistically significant, we lose the significance of the effect of income inequality. This is show numerically in Table 8 and Table 9 but it is far easier to see in the partial regression plots, Figure 6. It is clear to see that there is no relationship between the first-differences in crime rates and the first-differences of income inequality.

But, when we explore the effects on each type of crime, as shown in Table 8 and Table 9, we see a statistically significant positive relationship between property crime and income inequality. This relationship means that as the income of an area rises by £1,000[2] relative to its neighbours, ceteris paribus, the model predicts that 1 more crime will take place in that area. This result is statistically significant with a p-value of 0.032. However, the effect is very small. An area having a £1,000[2] greater average relative income will cost the average individual in that area 55p[2] in the additional cost of crime[3] [23]. This comes in the form of a single large cost for 1 household in the area, averaging around £2,200[2], or around 5% of household income, very inconvenient if you're the victim, but not revolutionary within the scale of the population. All of the other forms of crime have no significant relationship when observed on either the LA or MSOA scale except other crime with an equally small, but negative relationship.

However, when inequality is measured using the inequality of the house

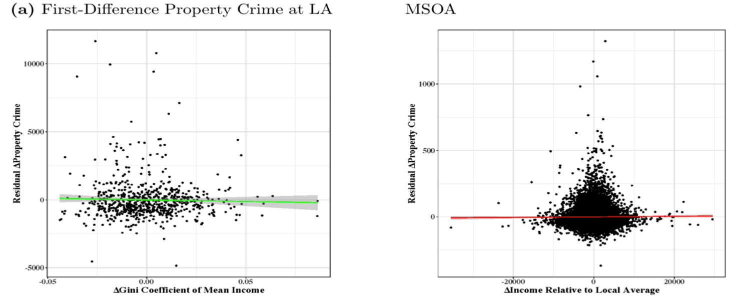
2 Monetary values adjusted to be given in terms of 2021 GBP.

3 Assumes a uniform increase of crimes within this group, the calculation for the average cost of a property crime is a weighted average of the costs of crimes within this group.

prices in that area, we see a strong positive relationship between the two variables. Table 10 shows the results of first-difference regressions using this generated Gini coefficient. For consistency, Figure 7 shows the results of this modelling, however, since the effects are small, they are hard to see graphically. This relationship suggests that the difference between perfect equality and perfect inequality across a population will be about 1.7 crimes per 100 people. A more realistic observation is that if inequality rises by one standard deviation, a rise in Gini coefficient of around 0.1, we can expect an increase of about 14 crimes in an area. However, the fit of this model is very weak, with the model explaining very little of the variation over time. This suggests that there are more significant causes of crime outside the model. While violent crime is not significant to the 95% confidence interval, its p-value of 0.064 is very close to the critical value.

Firstly, this shows that visible inequality has an effect on all three categories of crime. This means that, while the effect of one standard deviation is about the same for the sum of crime, since the effect is much smaller for property crime, and the effect it has on violent and other crime is greater, the average value of the crimes increases. This means that while one standard deviation of relative income can be said to cause an increase in the cost of crime of £3.30[2] per individual, an increase in Gini coefficient of one standard deviation, 0.1, leads to an increase in the cost of crime in the area of £8.82[2] per individual annually, or £70,530[2] per area[4].

Figure 6: These partial regression plots show the results of the first-differenced models of crime rates. The left column consists of LA aggregated data, the right shows MSOA aggregated data. The top row is property crime, the second row is violent crime, and the bottom is other crime.

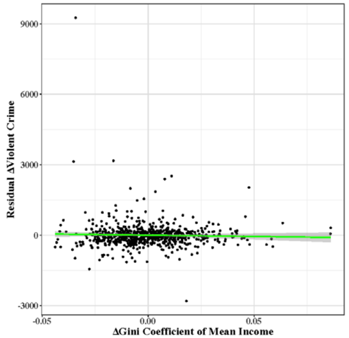


(a) First-Difference Property Crime at LA

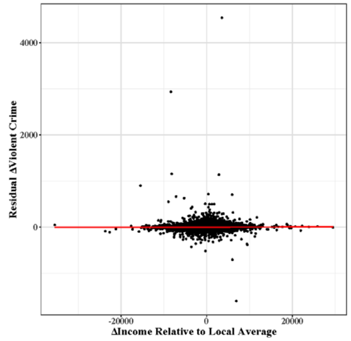
(b) First-Difference Property Crime at MSOA

4 This disparity is driven mostly by the far higher valuation of violent crimes, causing five times the cost of damages, calculated as before [23].

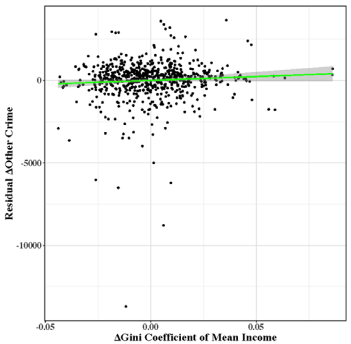
(c) First-Difference Violent Crime at LA



(d) First-Difference Violent Crime at MSOA



(e) First-Difference Other Crime at LA



24

(f) First-Difference Other Crime at MSOA

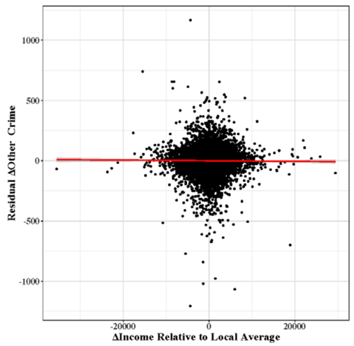


Table 8: Results of the regression on the first-difference model. Aggregated to the LA level, showing that there is not a significant causal relationship between income inequality and crime when observed at this level.

	Dependent variable:			
	Crimes	Property Crime	Violent Crime	Other Crime
Gini	1,152.414	−2,764.346	−1,191.566	5,108.326**
	(4,472.088)	(2,966.980)	(1,423.599)	(2,563.490)
Claimants (%)	−967.764*	−1,115.094***	−20.221	167.552
	(537.882)	(342.568)	(89.948)	(236.632)
Pop. Density	−2.710*	−3.725***	0.641**	0.373
	(1.476)	(1.175)	(0.317)	(0.520)
Population	0.370***	0.314***	0.047***	0.010
	(0.091)	(0.067)	(0.017)	(0.031)
Age 16–25 (%)	34.478	−84.471	−7.216	126.165
	(268.173)	(182.021)	(90.286)	(129.074)
Year 2014	−627.907	−956.649***	−124.022	452.764**
	(414.378)	(297.363)	(130.715)	(197.657)
Year 2016	−776.235***	−925.037***	−179.893**	328.695***
	(179.923)	(125.177)	(76.452)	(86.292)
Mean Income	−0.042	−0.012	−0.022	−0.007
	(0.033)	(0.021)	(0.022)	(0.020)
Observations	686	686	686	686
R^2	0.190	0.226	0.139	0.014
Adjusted R^2	0.181	0.218	0.130	0.004
F Statistic (df = 8; 678)	52.077***	99.998***	22.124***	6.788***
Note:				*p<0.1; **p<0.05; ***p<0.01

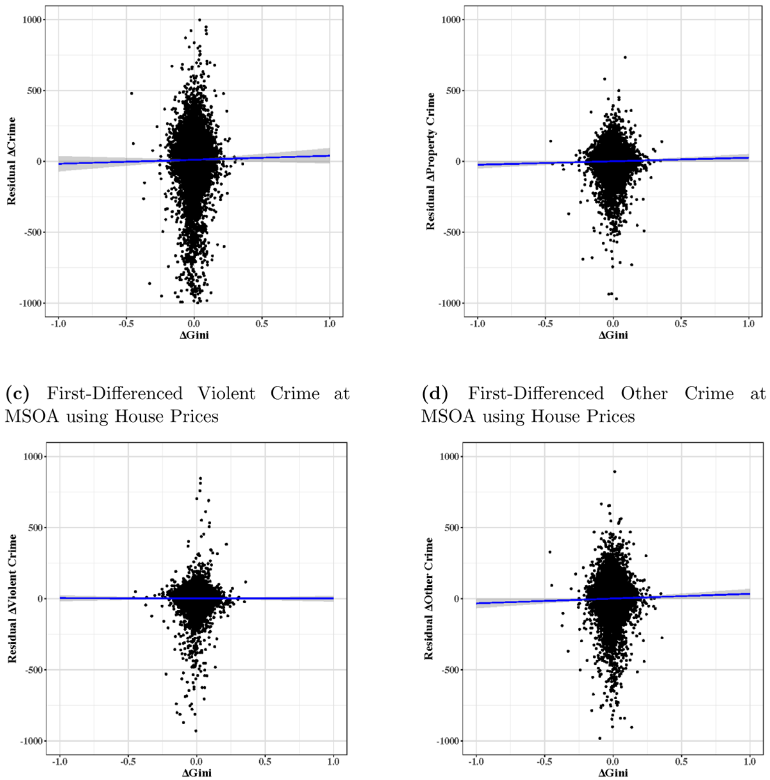
Table 9: Results of the regression on the first-difference model. Aggregated to the MSOA level, showing that there is possibly a relationship between localised income inequality and property crime rates, but with a very small effect.

	Dependent variable:			
	Crimes	Property Crime	Violent Crime	Other Crime
Income Relative to Neighbours	0.0003	0.001**	0.0005	−0.001*
	(0.001)	(0.0004)	(0.001)	(0.001)
Claimants (%)	−28.701***	−41.011***	−3.123**	15.432***
	(4.338)	(1.993)	(1.493)	(2.734)
Pop. Density	0.00001	−0.022***	0.013	0.008
	(0.015)	(0.006)	(0.009)	(0.007)
Population	0.116***	0.069***	0.022***	0.024***
	(0.014)	(0.010)	(0.007)	(0.006)
Age 16–25 (%)	−8.825***	−9.102***	−1.459	1.736
	(2.569)	(1.376)	(1.166)	(1.634)
Year 2014	−72.829***	−88.338***	−9.075***	24.584***
	(5.242)	(2.488)	(2.956)	(2.901)
Year 2016	−52.151***	−61.265***	−10.640***	19.754***
	(3.061)	(1.371)	(1.980)	(1.602)
Income	−0.001*	−0.001***	−0.001	0.0003
	(0.001)	(0.0003)	(0.001)	(0.0004)
Observations	13,152	13,152	13,152	13,152
R^2	0.049	0.110	0.019	0.012
Adjusted R^2	0.048	0.110	0.019	0.012
F Statistic (df = 8; 13144)	342.504***	1,597.816***	54.214***	60.622***
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 10: Results of the regressions on the first-difference model, using data aggregated at the MSOA level, with Gini coefficients are derived from house prices.

	Dependent variable:			
	Crimes	Property Crime	Violent Crime	Other Crime
Gini	135.473***	33.825***	53.268*	48.379**
	(45.508)	(12.831)	(28.775)	(19.230)
Income	−0.005***	−0.001*	−0.001	−0.004***
	(0.001)	(0.0003)	(0.001)	(0.0004)
Claimants (%)	−14.319*	−12.070***	−21.144***	18.895***
	(7.842)	(2.622)	(3.700)	(3.633)
Pop. Density	−0.088*	−0.052***	−0.079***	0.043**
	(0.045)	(0.013)	(0.023)	(0.018)
Age 16–25 (%)	6.638	−1.885	9.804***	−1.281
	(7.237)	(2.348)	(3.447)	(3.193)
Population	−0.132***	−0.022	−0.084***	−0.027
	(0.050)	(0.014)	(0.020)	(0.022)
Year 2018	23.773**	−16.662***	12.193*	28.242***
	(11.876)	(3.104)	(7.034)	(4.743)
Year 2019	7.765	−7.597**	11.258*	4.104
	(11.613)	(3.160)	(6.829)	(4.562)
Observations	13,073	13,073	13,073	13,073
R^2	0.011	0.024	0.029	0.073
Adjusted R^2	0.010	0.024	0.029	0.073
F Statistic (df = 8; 13065)	21.638***	40.841***	85.345***	141.013***
Note:	White [42]'s standard errors used; *p<0.1; **p<0.05; ***p<0.01			

Figure 7: This plot shows the results of the first-differenced model using inequality data from house pricing data. Each plot shows a different group of crimes: All crime (top left), property crime (top right), violent crime (bottom left), and other crime (bottom right).



(a) First-Differenced Total Crime at MSOA using House Prices

(b) First-Differenced Property Crime at MSOA using House Prices

(c) First-Differenced Violent Crime at MSOA using House Prices

(d) First-Differenced Other Crime at MSOA using House Prices

6 Conclusion

The findings of this paper are varied, depending on the measurement of inequality. When observed on the LA level, the presence of a correlation may simply be indicative of area fixed effects. This may be a relatively large amount of night-time entertainment in those areas, or increased gang presence within the more unequal authorities, unrelated to inequality. This makes causal interpretation from the cross-sectional analysis difficult. To remedy this, the data was differenced, removing the area fixed effects. Differencing this data removes the statistical significance of the relationship, suggesting that area fixed-effects are responsible for the cross-sectional correlation. This lack of a significant result persists, regardless of the type of crime. There is a positive relationship within other crime, but since this comes in a model with extremely weak fit, and costs of crimes in these group to society are low [23].

When considering the effects of being a rich area relative to one's neighbours, the opposite happens. The fixed effects here lead to a negative correlation. When using first-differenced data to remove these fixed effects, the only positive result came from increases in property crime. This makes theoretical sense as property crime offers the greatest economic benefits for criminals. However, since the effect is small, it is probably not economically significant.

Using the house price data makes this effect far clearer. In this case, the fixedeffects mask inequality leading to higher rates of crime. Since the unit social costs of violent crime are far higher, this effect is economically significant [23]. This suggests that the inequalities that cause higher rates of crime are not income inequalities, instead inequalities in tangible wealth. If individuals have a higher income, but put most of it into a savings account to save for retirement, then their possessions outside of these savings are not a higher value target for a criminal. This means that the use of identifiers such as a nice house or car in a driveway may be more useful for a criminal when trying to select a highly profitable target. While there may be effects of crime rates on house prices, which cause an endogeneity issue, this issue should be captured by the income measure being based on the mean price. The areas used within this study are relatively small, so if an area is labelled as a 'high crime area', we expect this to effect the prices of the whole area, not just the prices of some houses.

However, this model of crime had an extremely weak fit, with an adjusted R^2 of 0.010. This may suggest that economic indicators may not be the best variables to observe in order to predict changes in crime. These findings may also be susceptible to being driven by outliers. Criminology is heavily populated with papers presenting predictive crime analysis models using a variety of variables and algorithms, many of which fit their data far more closely than the model presented in this paper [25, 46, 22]. These suggest that the causes of crime may not be as simple as the basic economic motivations, other forces influencing the rates of crime in areas.

This paper advances upon previous findings, innovatively using house prices to capture smaller area wealth inequality data than has been found in the UK before in studies of a similar topic. This allows this research to bypass the problems experienced by Whitworth [43],and Metz and Burdina [28] who had to construct incomplete measures of inequality, as well as demonstrating that this method is key since there is too much information loss through the aggregation of area incomes to produce means. The findings using this data are statistically significant, but small. Further research using a dataset which spans a longer period of time would be beneficial, to ensure that this effect is long-term, and not a result of the years affected by the Covid-19 pandemic.

As with other recent works, this research does not account for the effects of security system installation [36, 15]. This has been found to be positively correlated with the rate of localised income inequality [12]. This suggests that the estimates made in this research are the lower bounds of the possible effects that income inequality has on crime rates as this countering pressure of buying protection is likely to be cancelling out some of the effects. The lack of positive findings in this field may come not from a lack of a relationship but because of a lack of easily accessible data [2]. In the process of this research multiple avenues to security spending data were explored with little success. This explains the findings being stronger when observing house prices as crime deterrence goods connected to the house should be factored into this price. These may include gates, fences, and security cameras.

Another shortcoming of this paper is that it cannot distinguish the mechanism through which this relationship arises. While the use of signals of income such as more expensive houses may be being used by criminals to target households for profit, we cannot rule out other explanations. Firstly, it may be the case that jealousy of these possessions is driving up crime rates in these areas. Blau and Blau [4] suggested that racial inequality leads to poorer inter-group relations, leading to an increase of inter-

racial crime. This same relationship may be occurring between the rich and poor, rather than criminals being profit motivated. Another possible explanation, given that the different effects were observed over two different periods of time, is that this relationship was simply a coincidence or that it was particularly strong between 2018 and 2021. Since this study is not able to tell these hypothesised mechanisms apart, no evidence of a mechanism can be taken. But we do conclude that there is weak evidence of a relationship.

7 References

[1] Ackerman, J. M. and D. K. Rossmo (2015). How far to travel? a multilevel analysis of the residence-to-crime distance. *Journal of Quantitative Criminology 31* (2), pp.

2] Ang, Y. Y. (2020). *'Rethinking Nine Big Questions' in China's Gilded Age: The Paradox of Economic Boom and Vast Corruption*, pp. pp. 180–212. Cambridge: Cambridge University Press.

[3] Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy* 76 (2), pp. 169–217.

[4] Blau, J. R. and P. M. Blau (1982). The cost of inequality: Metropolitan structure and violent crime. *American Sociological Review* 47 (1), pp. 114–129.

5] Boeckmann, R. J. and C. Turpin-Petrosino (2002). Understanding the harm of hate crime. *Journal of social issues 58* (2), pp. 207–225.

[6] Brush, J. (2007). Does income inequality lead to more crime? a comparison of crosssectional and time-series analyses of united states counties. *Economics Letters 96* (2), pp. 264–268.

[7] Buil-Gil, D., A. Moretti, and S. H. Langton (2021). The accuracy of crime statistics: assessing the impact of police data bias on geographic crime analysis. *Journal of Experimental Criminology 17* (1), pp. [Online].

[8] Cameron, S. (1988). The economics of crime deterrence: A survey of theory and evidence. *Kyklos 41* (2), pp. 301–323.

[9] Chintrakarn, P. and D. Herzer (2012). More inequality, more crime? a panel cointegration analysis for the united states. *Economics Letters 116* (3), pp. 389–391.

[10] Cohen, L. E. and M. Felson (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review 44* (4), pp. 588–608.

[11] Cohen, L. E., J. R. Kluegel, and K. C. Land (1981). Social inequality and predatory criminal victimization: An exposition and test of a formal theory. *American Sociological Review 46* (5), pp.505–524.

[12] Corvalan, A. and M. Pazzona (2022). Inequality, crime and private protection. *Economics Letters 210*, pp. 110–184.

[13] Danziger, S. and D. Wheeler (1975). The economics of crime: Punishment or income redistribution. *Review of Social Economy 33* (2), pp. 113–131.

[14] data.police.uk (2021). About data.police.uk.

[15] Demombynes, G. and B. Ozler (2005). Crime and local inequality in south africa. ¨ *Journal of Development Economics 76* (2), pp. 265–292.

[16] Department for Work and Pensions (2021). Relative low income.

[17] Doyle, J. M., E. Ahmed, and R. N. Horn (1999). The effects of labor markets and income inequality on crime: Evidence from panel data. *Southern Economic Journal 65* (4), pp. 717–738.

[18] Eberts, Paul Schwirian, K. P. (1967). Metropolitan crime rates and relative deprivation. *Criminologica 5* (4), pp. 43.

[19] Ehrlich, I. (1973). Participation in illegitimate activities: A theoretical and empirical investigation. *Journal of Political Economy 81* (3), pp. 521–565.

[20] Fajnzylber, P., D. Lederman, and N. Loayza (2002). Inequality and violent crime. *The Journal of Law and Economics 45* (1), pp. 1–39.

[21] Groff, E. R. and N. G. La Vigne (2002). Forecasting the future of predictive crime mapping. *Crime Prevention Studies 13* (1), pp. 29–58.

[22] H¨alterlein, J. (2021). Epistemologies of predictive policing: Mathematical social science, social physics and machine learning. *Big Data & Society 8* (1).

[23] Heeks, M., S. Reed, M. Tafsiri, and S. Prince (2018). The social and economic costs of crime.

[24] Hooghe, M., B. Vanhoutte, W. Hardyns, and T. Bircan (2011). Unemployment, inequality, poverty and crime: spatial distribution patterns of criminal acts in belgium, 2001–06. T*he British Journal of Criminology 51* (1), pp. 1–20.

[25] Joshi, C., S. Curtis-Ham, C. D'Ath, and D. Searle (2021). Considerations for developing predictive spatial models of crime and new methods for measuring their accuracy. *ISPRS International Journal of Geo-Information 10* (9).

[26] Kelly, M. (2000). Inequality and crime. *The Review of Economics and Statistics 82* (4), pp. 530–539.

[27] Levitt, S. D. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *The American Economic Review 87* (3), pp. 270–290.

[28] Metz, N. and M. Burdina (2018). Neighbourhood income inequality and property crime. *Urban Studies 55* (1), pp. 133–150.

[29] Neumayer, E. et al. (2004). Is inequality really a major cause of violent crime? evidence from a cross-national panel of robbery and violent theft rates. Law and Economics [online] (0312002). URL: *https://www.researchgate.net/profile/EricNeumayer/publication/23746976IsInequalityreallyaMajorCauseofViolentCrimeEvidenceFromaCrosNationalPanelofRobberyandViolentTheftRates/links/5be890cc92851c6b27b837dd/Is–Inequality–really–a–Major–Cause–of –Violent–Crime–Evidence–From–a–Cross – National – Panel – of – Robbery – and – Violent – Theft – Rates.pdf.*

[30] Office for National Statistics (UK) (2020). Income estimates for small areas, england and wales.

[31] Office for National Statistics (UK) (2021). Lower layer super output area population estimates.

[32] ONS Geography (2020). Middle layer super output areas (december 2011) boundaries full clipped (bfc) ew v3.

[33] Papaioannou, K. J. (2017, 01). "Hunger makes a thief of any man": Poverty and crime in British colonial Asia. *European Review of Economic History 21* (1), pp.1–28.

[34] Sampson, R. J. and W. B. Groves (1989). Community structure and crime: Testing social-disorganization theory. *American journal of sociology 94* (4), pp. 774–802.

[35] Saridakis, G. (2004). Violent crime in the united states of america: A time-series analysis between 1960–2000. *European Journal of Law and Economics 18* (2), pp. 203–221.

[36] Scorzafave, L. G. and M. K. Soares (2009). Income inequality and pecuniary crimes. *Economics Letters 104* (1), pp. 40–42.

[37] Shaw, C. R. and H. D. McKay (1942). *Juvenile delinquency and urban areas*. Chicago: University of Chicago press.

[38] Soares, R. R. (2004). Development, crime and punishment: accounting for the international differences in crime rates. *Journal of Development Economics 73* (1), pp.155–184.

[39] Stack, S. (1984). Income inequality and property crime. *Criminology 22* (2), pp.229–256.

[40] Stucky, T. D., S. B. Payton, and J. R. Ottensmann (2016). Intra- and interneighborhood income inequality and crime. *Journal of Crime and Justice 39* (3), pp.345–362.

[41] Velleman, P. F. and R. E. Welsch (1981). Efficient computing of regression diagnostics. *The American Statistician 35* (4), pp.234–242.

[42] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48* (4), 817–838.

[43] Whitworth, A. (2013). Local inequality and crime: Exploring how variation in the scale of inequality measures affects relationships between inequality and crime. *Urban Studies 50* (4), pp. 725–741.

[44] Williams, K. R. (1984). Economic sources of homicide: Reestimating the effects of poverty and inequality. *American Sociological Review 49* (2), pp. 283–289.

[45] Wilson, J. Q. and B. Boland (1978). The effect of the police on crime. *Law and Society Review 12* (3), pp. 367–390.

[46] Zhang, X., L. Liu, L. Xiao, and J. Ji (2020). Comparison of machine learning algorithms for predicting crime hotspots. *IEEE Access 8*, pp. 181302–181310.

[47] Zoopla Limited © (2022). Zoopla property data. Economic and Social Research Council. 2022 [data collection]. University of Glasgow - Urban Big Data Centre.

Is It the Woman's Call? Women's Mobile Phone Access and Household Consumption in Rural India

May 2, 2022

Abstract

Mobile phones facilitate women's social networks, increase access to new informa tion, and substitute for transportation because of the security risk of traveling alone in the rural developing world. To examine how mobile phones affect household decision-making, this paper asks how the consumption behaviors differ for house holds with at least one woman accessing a mobile phone as opposed to households with no women accessing a mobile phone by incorporating the collective household bargaining model. A household's logged total consumption and consumption shares are calculated using the India Human Development Survey of 2011. Using Ordinary Least Squares regression analysis and a Two Stage Least Squares instrumental vari ables design, this paper finds that overall consumption is 10% lower for households with at least one woman accessing a mobile phone. These households with women's mobile phone access were also found to spend more on food and household financ ing and less on transportation. These results imply a stronger preference for savings and food spending in households with higher female bargaining power. Further em pirical research could investigate whether these changes are persistent over time in rural India. This paper contributes to the literature connecting mobile phones and international development to the empirical research on household bargaining in the developing world.

Content

1 Introduction

This paper considers how rural Indian women's mobile phone access can change how her household spends money. India has one of the widest gender gaps in mobile phone ownership and access in the world. The Financial Inclusion Insights 2016 data report that mobile phone access differs by 15% and phone ownership by 45% (Barboni et al. 2018). Even when controlling for education, age, state, marital status, rural or urban location, and poverty status, substantial gender gaps exist with a minimum of 10% (Barboni et al. 2018). Women's ownership and access of mobile phones has grown in the past few years such that investigating its impacts is more important now than ever before (GSMA 2021). Gaps in access to basic phone functions like sending and receiving calls are smaller, providing some optimism for a closing gender gap that can open new possibilities for these women (Moore and Sabherwal 2017).

Significant structural inequalities in India keep women from accessing mobile phones and using them productively once acquired. Marked gender gaps exist in education, English literacy, and access to information channels like news and television (Scott et al. 2021; Potnis 2016; Jyoti 2021; Rink et al. 2021). These gaps significantly stunt the financial and digital literacy of poor women in particular, then making them uncomfortable with and disinterested in an unfamiliar product that could disrupt their lives (Rink et al. 2021). There also exist significant cultural taboos keeping rural, poor Indian women from accessing phones at all. Conservative, rural communities maintain restrictive norms about women's access to phones. Single women are often harassed through mobile phones, so conservative communities stigmatize women from using mobiles and being exposed to such threats (Pande and Schaner 2017). Additionally, single women who do own phones are seen as being promiscuous and violating norms of courtship and family honor (Pande and Schaner 2017; Sonne 2020; Barboni et al. 2018).

Gender norms limit married Indian women's access as well. The limits to independent finances and digital literacy mean that Indian women's use of mobile phones can greatly depend on the ownership and generosity of their husbands (Scott et al. 2021; Moore and Sabherwal 2017; Pande and Schaner 2017; Barboni et al. 2018; Group 2016). Phone access may conflict with traditional gender roles: "Norms dictate that a woman's primary responsibility is to take care of her family and household. This home-centric role leaves women with few opportunities to use the phone for socially-acceptable and productive purposes" (Barboni et al. 2018).

These systemic and cultural barriers can block rural Indian women from a range of benefits mobile phones can bring to their lives. The relationship between rural Indian women's mobile phone access and her relative power in the household can be described by one survey respondent's experience: "A phone in this area of India is more than a device. It is a powerful symbol – of pride, empowerment and independence" (Group 2016).

The barriers that rural Indian women face in accessing mobile phones have been thoroughly researched, but as the technological gender gap closes, research should pivot the equally important question of how mobile phone access affects their lives. Consumption outcomes are one way in which the empirical literature on household welfare in the developing world has investigated this question. Even a cursory understanding of the literature on development economics would reveal that a household's investments in education, nutrition, financing, and health have profound effects on things like inter-generational poverty, local markets, and economic development. The household is the base unit of a community, so understanding the forces that affect its consumption outcomes can have far-reaching implications for different types of policy.

This paper considers a difference in means and ordinary least squares regression methods, ultimately implementing a two-stage least squares instrumental variables design to account for the endogeneity between women's mobile phone access and how households allocate their consumption spending. Two sets of instrumental variables are considered, one set being village-level characteristics that affect general mobile phone uptake and household bargaining characteristics of families that affect the mobile phone access of women in those households. This paper finds that households with at least one woman accessing a mobile phone have lower levels of consumption but have higher consumption shares dedicated to food and nutrition. One weakness of this paper is that the village level instruments used in the consumption share regressions instrument for overall mobile phone uptake rather than that of women. Perhaps more specific instruments could reveal changes in the other consumption shares with greater statistical precision.

This paper begins with a brief literature review of the effects of mobile phones on people in the developing world and women's mobile phone access in rural India in particular. It then discusses the use of the household bargaining model to explain how mobile phone use may or may not affect total household consumption or how household consumption is allocated across ten spending categories. This paper then employs a two-stage least squares instrumental variables design to examine how changes in household bargaining power through women's mobile phone access affected these household consumption outcomes. This paper discusses the results of these regressions before drawing conclusions in the final section.

2 Literature Review

2.1 *Mobile Phones in the Developing World*

Research on mobile phones in the developing world has found a wide range of benefits. Much recent literature focuses on the impacts of mobile phones or mobile money on women's autonomy indicators or a range of health outcomes (Rotondi et al. 2020; Mo han et al. 2020; Kruse et al. 2019; Biswas et al. 2021; Stark 2020). Some have found that mobile phones lower information costs for firms, increase cooperation, and decrease price dispersion in local markets (Jensen 2007; Aker and Mbiti 2010; Goyal 2008). The social networks enabled by mobile phones can help disperse information, opportunity, and assets for households. The reduction in travel and information costs has been one of the most significant benefits of mobile phone use for households, individuals, and firms alike. Time poverty, or the limited time one has for economically productive activities, plagues women in the developing world (Hyde, Greene, and Darmstadt 2020). Mobile phones can share information and resources with women and households that would otherwise take much time, money, and distance to acquire (Muto and Yamano 2009; Aker and Mbiti 2010; Jensen 2007; Overa 2006). For example, Goodman 2005 found that mobile phones enabled households to maintain social networks that, in turn, shared information and economic opportunity. Samuel et al. (2005) similarly found that the types of information spread to households were business information, friends and family contacts, and job prospects.

A strand of literature connects access to mobile phones in the developing world to household economic welfare. Adoption of mobile phones and mobile money has been found to increase the level and growth rate of household per capita consumption, level of overall household expenditures, and household real consumption while lowering poverty incidence across many developing countries (Munyegera and Matsumoto 2016; Labonne and Chase 2009; Tankari 2018; Beuermann, McKelvey, and Vakis 2012). Other papers have found that mobile phones enable risk sharing across social networks, increased financial resiliency, and long-term household consumption increases with greater effects for female-headed households (Jack and Suri 2014; Rettie, 2008; Goodman 2005; Duncombe 2014; Riley 2018; Suri and Jack 2016; Sekabira and Qaim 2017). Many previous studies did not disaggregate their findings by gender, however. Some that do suggest that while per capita household consumption increased with mobile phone ownership, there was not a statistically significant difference in effects for male versus female mobile phone owners (Tankari 2018).

A small group of literature has examined the uses and effects of mobile phones on the lives of rural Indian women specifically. The primary use of mobile phones was to maintain social connections with friends and family, particularly during emergencies (Souter et al. 2005; B. S. Mehta 2013; B. S. Mehta and N. Mehta 2014; Malhotra, Kanesathasan, and Patel 2012; Group 2016). Even illiterate women find that audio or picture messages still enable efficient communication, saving on monetary and time costs of transportation (Moore and Sabherwal 2017; B. S. Mehta 2013; Malhotra, Kanesathasan, and Patel 2012). Indeed, substituting mobile phone communication for transportation made women value mobile phones more as a savings tool than an income generating tool (Souter et al. 2005). Phones even make women feel safer while travelling, a great safety concern for rural Indian women (Moore and Sabherwal 2017; Group 2016; Malhotra, Kanesathasan, and Patel 2012). In their survey of 200 rural Indian women, Mehta and Mehta (2014) summarize their varied findings: "The majority of mobile phone users had perceived very high impact on their personal relationship, education and health followed by moderate impact on social networking, economic opportunities and higher autonomy in decision-making." Improvements in autonomy were seen through enhancing "the ability of women to plan, coordinate and search for better information in comparison to non-users" (B. S. Mehta and N. Mehta 2014). Malhotra et al. (2012) find that these changes to women's activities can shift perceptions about women's roles and positions in the family and society. They also find that the benefits of increased information gathering disseminate into their community and help non-users as well (Mehta and Mehta 2014). One survey respondent summarizes succinctly what mobile phone access has done for her and her community: "The sense of independence and confidence is unmatchable. Maybe this is the first step towards a larger social change" (Group 2016).

Some literature recognizes that mobile access does not immediately translate to change. Women without access may not seek it out because do not know the potential benefits (Moore and Sabherwal 2017; Malhotra, Kanesathasan, and Patel 2012; Sonne 2020). For example, Souter et al. (2005) finds that non-users consider mobile phones unimportant for accessing information while Mehta and Mehta (2014) find that users primarily valued mobile phones for increasing access to information about education, health, and personal relationships. Even when the benefits are known, men can act as gatekeepers to the ad vanced functionality of phones like financial transactions or information gathering (Moore and Sabherwal 2017; Pande and Schaner 2017; Group 2016; Sonne 2020). Different uses for mobile phones were also valued by different women depending on socioeconomic characteristics, regardless of family makeup. Souter et al. (2005) found that richer and more educated populations valued mobile phones for financial activity while rural, uneducated populations sometimes found phones unhelpful for traditional means of economic activity done face-to-face. These conflicting interests and conditions in rural India create an interesting case study for examining the intersection of technology, women's economic activity, and international development.

2.2 Household Bargaining Model

The collective household bargaining model could bring together the literature on gender, mobile phones, and household consumption allocations. The classic collective model of household decision making suggests that decisions are not completed unless the utility of each spouse is greater than the utility of an "outside option" at their individual dispos als (Chiappori 1988, 1992; Bourguignon et al. 1993). A simple utility equation can be maximized with respect to the household's budget constraint and access to public goods *Q*:

$$U_1(C_1,Q) + Z \times U_2(C_2,Q)$$

The first unit in the utility equation refers to the primary decision maker, likely the household head, while the second unit in the utility equation refers to the "secondary" actor with a smaller weight given to their preferences. Z is the relative weight of the second spouse's utility determined by "distribution factors" of the environment independent of the budget constraint that constitute relative standing in the household (e.g. relative income, legal standing, proportion of women in the population). Z takes a value between 0 and 1, referring to the weight given to this actor's preference in proportion to that of the primary actor. Maximizing this

equation finds the local maxima of the utility function subject to two conditions:

$$Max_{C1,\,C2}\, U_1(C_1,Q) \text{ s.t.}$$
$$U_2(C_2,Q) \geq \bar{U}\,(C_2,Q)$$
$$C_1 + C_2 = Y$$

That utility maximization is equivalent to the one below due to Z acting as the Lagrange multiplier that makes these conditions hold.

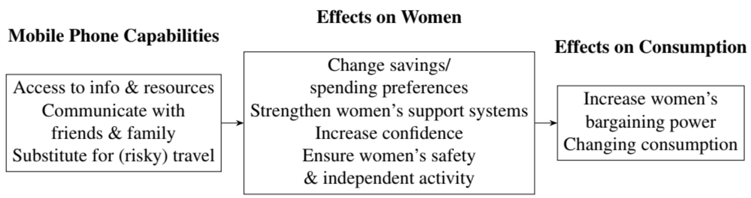
$$Max_{C1}\, U_1(C_1,Q) + Z \times U_2(Y - C_1,Q)$$

This utility maximization is Pareto-efficient, meaning that if one spouse receives greater utility from an outcome, the utility of the other spouse decreases (summarized by Baland and Ziparo 2018). As mentioned, this multiplier refers to the distribution factors that would make one actor's preferences relatively less important than the other's.

In their survey of household bargaining literature, Baland and Ziparo (2018) find that "distribution factors" are particularly important in the developing world. For example, a change in the relative incomes of spouses affects the patterns of household expenditures and decision making (e.g. Duflo 2003; Haddad and Hoddinot 1995; Thomas 1994). They show: "In the context of the collective model, this change is directly interpreted as a change in the outside option of the spouse, which directly affects his/her bargaining power, and therefore his/her Pareto weight." Papers like Suri and Jack (2016) implicitly rely on the unitary model, a model long since rejected, when they state that expenditures change to become more efficient because Pareto weights make even one-sided spousal decisions efficient according to these changing "distributional factors."

Mobile phones could present an increase in women's Pareto weight by granting them greater outside options. This can be through a variety of mechanisms: greater support networks through community, access to finance and enterprise, or greater information that changes expectations about spousal decision making. This could entail greater household expenditures on goods more likely to be preferred by these women (e.g. child's education, nutrition, healthcare). Some literature suggests that consumption of male-favored goods like intoxicants and processed foods decreases once women have greater power over household consumption decisions while food and education consumption increases (Fafchamps, Kebede, and Quisumbing 2009; Wang 2014; Pajaron 2011; Duflo and Udry 2004).

Synthesizing this literature, the following flow chart details the hypothesized effects of mobile phones on women's preferences, self-perceptions, social networks, and range of activity.



One issue with this model is that a woman's access to mobile phones could be a result of collective bargaining. For those households where women have great enough Pareto weight to bargain for cell phones, they likely bargain for other consumption changes as well. This informs the empirical strategy of this paper by changing which instrumental variables are used for which type of consumption variable identified. Even with these empirical modifications, this paper suggests that these theoretical considerations must be taken on the empirical household bargaining literature and must contribute understanding from sociology, philosophy, and cultural studies in addition to empirical disciplines. In general, the collective model of household bargaining is a relevant and multifaceted perspective to contribute to this paper.

I hypothesize that household bargaining has a more direct effect on how household consumption is allocated rather than on total household consumption. Advanced phone functions that could directly provide tools for saving or income generation like mobile money have very low adoption rates and are likely out of reach for this demographic due to wide financial and technological literacy gaps by gender (Barboni et al. 2018; Mohan et al. 2020). These considerations suggest that mobile phones may have a direct effect and/or an indirect effect on household consumption. Their "direct" effects may come through the first two stages of the flow chart above: changing preferences and increasing women's scope of activity through greater security and independence. These may translate into the "indirect" effects on their household bargaining power in the third stage of the flow chart: women may take these changes due to mobile phones and advocate for their preferences from a stronger bargaining standpoint.

This paper seeks to make a number of contributions to these strands of literature. First, this paper contributes to the scant-yet-growing number of papers investigating the link between mobile phones and household consumption in the developing world. This paper is also unique in that it investigates how *women's* mobile phone access affects these outcomes by incorporating a household bargaining model of gender and power dynamics. Lastly, the field has been greatly saturated with cross-country analyses or individual studies of African countries. With a focus on India, this paper will contribute to the growing literature on mobile phones, household welfare, and gender in South Asia.

3 Data Source

This paper uses the 2011-2012 wave of the Indian Human Development Survey (IHDS II). The IHDS II consists of cross-sectional survey data taken from eligible women (those ever-married women ages 15-49), their corresponding households, and their villages. This nationally representative, multi-topic survey has covered 42,152 households in 1,420 villages and 1,042 urban neighborhoods across India. With a recontact rate of 85%, the IHDS II consists mostly of re-interviewed households from the first survey wave of 2004- 2005 (IHDS I). All states and union territories of India are covered in this survey with the exception of Andaman & Nicobar and Lakshadweep.

While the IHDS covers over 150,000 individuals, this paper uses a rural, household level sub sample. First, households that lived in any urban area according to the 2011 Indian Census were dropped, creating a solely-rural sample of "more developed villages" and "less developed villages". Isolating a primarily low-income, rural, and less-developed sample makes this investigation more interesting and robust. This study focuses on a demographic for whom mobile phones are more difficult to access and could improve their livelihoods greatly. It is also known that urban individuals are more likely to own mobile phones partly due to higher incomes, greater wealth, and network effects of urban living. By dropping the vast majority of these individuals from the sample, there is less endogeneity for this paper's model to overcome. Only households interviewed using the Village, Household, and Eligible Women surveys are kept. Then, individual responses from the Eligible Women survey were used to create the following household-level variables: any woman accessing a phone, the number of women accessing a phone, female headed households, any woman earning an income outside the home, number of working women, women's labor force participation rate for the household, difference in highest male education and highest female education, ratio of under 14 children to adults, whether a woman's name is on the household papers, whether the female respondent or a senior female are involved in purchasing decisions, and region. The resulting data set is 16,447 unique households. For summary statistics and regressions, sample weights from the Eligible Women survey are used and standard errors were clustered at the primary sampling unit (PSU) level.

Next, I will describe some basic summary statistics for the sample. Table 1 computes basic summary statistics on the households in the final data set using the weights available from the Eligible Women data set. It assembles the mean, standard deviation, minimum and maximum values for the following variables: household size, ratio of children to adults, highest male education, highest female education, index of total household assets (0-33), household total income, and ratio of expenditures per capita with the 2011 poverty line.

Table 1: Summary Statistics: Household Characteristics

	Mean	SD	Min	Max
Any Woman with Mobile Access	.5013982	.5000132	0	1
Household Owns Cell Phone	.8408511	.365826	0	1
Size of Household	5.185546	2.289496	1	33
Household Ratio of Adults:Children	.5967419	.6473904	0	6
Highest Male Adult Education	7.414098	4.883595	0	16
Highest Female Adult Education	4.862854	4.818994	0	16
Total Household Assets (Index 0-33)	14.2743	5.291093	1	31
Household Total Income	105486.5	198455.9	100	1.12e+07
Expenditures Per Capita Below 2011 Poverty Line	.1757719	.3806376	0	1
Observations	16450			

Table 2 contains summary statistics on village characteristics and household bargain ing variables. 100% of the villages surveyed had mobile phone service available, indicating that no households are prohibited from mobile phone access due to infrastructural issues. Other village characteristics in the table are: how many years a village has had mobile phone service, average hours of electricity per day, distance to the nearest bus availability, and distance to the district headquarters. The household bargaining variables summarized refer to traits that the literature identifies as critical to women's bargaining, including a woman's education, employment status, and economic status before marriage (Baland and Ziparo 2018). These include: difference in highest male and female education, the relative economic status of a woman's natal family to her marital family, whether a woman's name is on the household papers, and whether any woman works outside the home.

Table 2: Summary Statistics: Village and Household Bargaining Characteristics

	Mean	SD	Min	Max
Village: Years Had Mobile Service	7.027579	3.631385	1	85
Village: Electricity Hours per Day	13.4515	6.526496	1	24
Village: Closest Bus Availability	2.028602	3.83924	0	30
Highest Male Educ - Highest Female Educ	2.50474	4.876639	-16	16
Women's Natal Family v. Marital Family Status	1.362586	.6507673	1	3
Any Female Name on Household Papers	.1713859	.3768573	0	1
Any Woman Works Outside Home	.2713796	.444685	0	1
Observations	16450			

Table 3 presents summary statistics on women's mobile phone access in the house hold, the household's total consumption and per capita consumption (2011 Rupees) as well as consumption shares of ten categories of consumption. These were calculated by multiplying monthly expenditures by 12, summing relevant expenditures by category, and dividing them by the annual total consumption per household. These categories are listed and described in Appendix 1. These categories are: food and nutrition, education, finances, transportation, energy and utilities, household goods, jewelry and ornaments, discretionary spending, care, and "vices". Doss 2005 enumerates these categories along similar lines. The elements of these categories are listed in Appendix 1. Some reported consumption expenditures common day-to-day, like food and household goods, were recorded on a monthly basis while larger expenses, like education or discretionary spending, were recorded on an annual basis. Total consumption per household was recorded annually, so monthly consumption expenditures were transformed into annual measures before each share was computed. Jewelry and ornaments was a consumption category left as its own because some literature cites jewelry and gold as an acceptable form of women-owned property and stores of value in India (Mehrotra 2004). The "care" category included goods and services that relate to a woman's public responsibilities to her family such as medicine, clothing, and cleaning supplies. Lastly, the category of "vices" contains consumption of intoxicants and processed foods. Some literature suggests that these are male-favored goods whose consumption decreases once women have greater power over household consumption decisions (Wang 2014; Pajaron 2011; Duflo and Udry 2004).

Table 3: Summary Statistics: Consumption and Shares

	Mean	SD	Min	Max
Total HH Consumption	106383.8	104139.6	2832	4028836
Consumption Per Capita (2011 Rupees)	22700.78	25375.81	708	1007209
Food	.4650041	.1465841	0	.9165302
Education	.0444625	.0724993	0	.8336733
Finances	.0200587	.0475555	0	.9433836
Transportation	.06282	.0740986	0	.7899154
Energy	.0914487	.0533613	0	1
Household Goods	.0354695	.0600662	0	.87564
Jewelry	.0106158	.0525282	0	.8311032
Discretionary	.0573356	.0934179	0	.9032336
Care Goods	.1837542	.1257027	0	.9937907
Intoxicants and Processed Foods	.029031	.0320692	0	.5230408
Observations	16447			

Statistics on total household consumption and per capita consumption are in 2011 Rupees.
All other statistics are shares of household annual consumption.

Consumption Group	Elements
Food and Nutrition	edible oils and vanaspati, gue and sweets, meat, pulses, cereals, other cereals, rice, fruits and nuts, tea and coffee, salt and spices, vegetables, milk, milk products, eggs, sugar, wheat or flour
Education	schoolbooks, private tuition fees, school and college fees
Finances	insurance premiums, consumer taxes, house rent, other rents (e.g. appliances), house loan installment
Transportation	personal transportation equipment, petrol or vehicle maintenance, transportation
Energy and Utilities	kerosene, telephone/cable/internet charges, household electricity, household fuel
Household Goods	repair and maintenance, other personal goods, personal care and household items, cooking and household appliances, crockery and utensils, furniture and fixtures, domestic services (servants, barber, laundry), household items
Jewelry and Ornaments	
Discretionary Spending	social functions, vacations or holidays, recreational goods, entertainment, restaurants and eating out
Care	footwear, clothing and bedding, in- or out-patient medical services, soap, toiletries, therapeutic appliances (eyeglasses, hearing aids)
Vices	paan/tobacco/intoxicants, processed foods

Table 4 reports the difference in means of total consumption, per capita consumption, and consumption shares between households with and without at least one woman accessing a mobile phone. There are statistically significant differences in all allocation shares and total consumption between the two groups, but their magnitudes are quite small: most differences range between 0.2 and 3 percentage points. Statistically significant differences in these means still suggest that this question should be investigated with greater statistical rigor.

Table 4: Difference in Means of Household Consumption Variables

	HHs with Fem Access		HHs Without Fem Access		Difference	
	Mean	SD	Mean	SD	Difference	T Stat
Total HH Consumption	128545.313	129837.749	99334.748	94809.029	-29210.564***	(-16.484)
Logged Total Consumption	11.525	0.640	11.286	0.629	-0.239***	(-24.177)
Consumption Per Capita (2011 Rupees)	25931.649	28033.067	20828.417	21360.225	-5103.232***	(-13.136)
Logged Per Capita Consumption	9.919	0.636	9.727	0.611	-0.192***	(-19.694)
Food	0.447	0.146	0.481	0.144	0.034***	(14.924)
Education	0.052	0.076	0.036	0.067	-0.016***	(-14.126)
Finances	0.026	0.056	0.016	0.043	-0.009***	(-12.022)
Transportation	0.071	0.081	0.060	0.070	-0.011***	(-9.183)
Energy	0.093	0.051	0.095	0.056	0.002*	(2.560)
Household Goods	0.041	0.075	0.034	0.059	-0.007***	(-6.241)
Jewelry	0.014	0.060	0.009	0.050	-0.005***	(-5.262)
Discretionary	0.058	0.091	0.054	0.087	-0.004**	(-2.635)
Care Goods	0.172	0.121	0.181	0.124	0.009***	(4.596)
Intoxicants and Processed Foods	0.026	0.028	0.033	0.035	0.006***	(12.447)
Observations	8247		8200		16447	

Statistics on total household consumption and per capita consumption are in 2011 Rupees.
All other statistics below those are shares of household annual consumption.

Much literature has shown that women that are more likely to have access to mobile phones are younger, more educated, living in urban areas, and have higher income (e.g. Roessler et al. 2018; Barboni et al. 2018; Potnis 2016). Indeed, this trend remains true for this sample. Table 5 displays the difference in means on household demographics. It indicates that households with a woman accessing a mobile phone are wealthier, have higher total income, greater education, and less of a difference between highest male and female education. Any one or combination of those variables could affect a woman's mobile access and household consumption, as has been found in many papers (e.g. Roessler et al. 2018; Barboni et al. 2018; Potnis 2016).

Table 6 displays the difference in means on household bargaining characteristics as suggested by Mohapatra and Simon 2017. Using the 2005 wave of the IHDS, they ask how women's intra-household bargaining power affects the adoption of improved cook stoves. Many of the variables used in this paper and assessed in Table 6 are factors identified in the intra-household bargaining literature of the developing world as affecting consumption decisions. The table suggests that for households with at least one woman accessing a mobile phone: the economic status of women's natal families to the marital families is more equal, they are more likely to have a woman's name on the house legal papers, and they are more likely to be headed by women. There was an unexpected difference identified by Table 6: households with no female mobile access had more women working outside the home. It was hypothesized that women earning income from outside the home would have more relative bargaining power in deciding how total household income was spent. An explanation of this situation could relate to expectations of women's labor differing across socioeconomic statuses, something that could be investigated in further research. Even OLS estimates may be biased because household bargaining characteristics affect both a woman's mobile phone access *and* her household's consumption patterns. This and further considerations will be discussed in the evolving empirical strategy.

Table 5: Difference in Means of Household Characteristics

	HHs with Fem Access		HHs Without Fem Access		Difference	
	Mean	SD	Mean	SD	Difference	T Stat
Total Household Assets (Index 0-33)	16.488	5.220	13.232	5.280	-3.257***	(-39.778)
Household Total Income	140602.995	248737.867	95069.708	135180.986	-45533.286***	(-14.598)
Size of Household	5.427	2.357	5.205	2.308	-0.222***	(-6.092)
Highest Adult Education in a Household	9.128	4.469	7.034	4.754	-2.094***	(-29.097)
Highest Male Educ - Highest Female Educ	1.931	4.864	2.949	4.729	1.018***	(13.200)
Household Ratio of Adults:Children	0.593	0.616	0.597	0.621	0.004	(0.420)
Observations	8248		8202		16450	

Table 6: Difference in Means of Household Bargaining Characteristics

	HHs with Fem Access		HHs Without Fem Access		Difference	
	Mean	SD	Mean	SD	Difference	T Stat
Women's Natal Family v. Marital Family Status	1.370	0.660	1.346	0.637	-0.024*	(-2.386)
Any Female Name on Household Papers	0.185	0.388	0.148	0.355	-0.037***	(-6.296)
Female Household Head	0.149	0.356	0.093	0.291	-0.056***	(-11.062)
Any Woman Works Outside Home	0.236	0.425	0.266	0.442	0.030***	(4.452)
Household Female Employment Rate	0.195	0.399	0.239	0.438	0.044***	(6.744)
Observations	8248		8202		16450	

4 Empirical Strategy

The empirical strategy of this paper evolves due to a number of considerations. The differences in means in the previous section show that there is some statistically significant difference in consumption outcomes for households with and without female mobile phone access. The literature review suggests three ways mobile phones can affect women with access

to them and, thus, the spending outcomes of the households they are in. These three channels are facilitating contact with friends and family, substituting for risky solo travel, and increasing access to new information and resources that affect the types of activity she is capable of. The other characteristics that affect consumption decisions are household and village characteristics. Thus, the difference in means does not sufficiently account for these characteristics. This paper first tries an Ordinary Least Squares regression analysis of the logged total consumption, logged per capita consumption, or individual consumption shares on whether the household had at least one woman accessing a mobile phone and a range of controls. The equation for the OLS strategy is below.

This paper first presents simple OLS regressions of each total consumption or consumption share variable regressed on a dummy indicating whether a household had any woman accessing a mobile phone and a panel of controls. This econometric model is specified as:

$$Consumption_i = \alpha_0 + \alpha_1 Mobile_i + \alpha_2 \boldsymbol{V}_i + \varepsilon_i$$

where α_1 is the coefficient of interest, $\boldsymbol{V}_i$ is a panel of control variables selected, and ε is the error term. $Consumption_i$ as a dependent variable represents one of the ten consump tion shares, ranging from 0 to 1, the log of the household's total consumption, and the log of the household's per capita consumption.

A range of control variables are included in different groups relating to different types of characteristics that may influence the results. The first set is a range of household characteristics: an index of wealth, highest adult education in the household, the ratio of the number of adults to children in the household, number of people in the household, and the log of the household's total income. The second set of controls are characteristics of the village that would affect the household's employment, could change the support systems available to a woman, and may indicate what infrastructural resources are available to the household. These include: electricity hours per day, the presence of employment guarantee programs (MG-NREGA, Food for Work, or others), and the presence of a Mahila Mandal (a women's association akin to a self-help group). The last set of controls are variables that can be both indicators and causes of women's increased household bargaining power. The final control variable included is an indicator for whether the household owns a cellphone. Including this variable lets the regressions control for the cellphone access of men. The household questionnaire from which this vari able is taken is given to the head of the household who is most likely male (88% of the households in the sample have a male head). Thus, household ownership of a cell phone and male access (for 88% of the sample) are controlled for by this variable. This isolates the effects that women's mobile phone access may have on consumption away from the effects of a household's ownership of a cellphone.

In the framework of the household bargaining problem, potential endogeneity should be addressed. It is theorized in this paper that there are variables in the household bargaining system that are *both* influential in the household bargaining model, and are influenced by other variables in the model. For example, the household bargaining literature mentioned previously describes the likelihood of a woman working outside the home as a determining characteristic of capabilities like access to a mobile phone, but it is also described as an outcome of the confluence of a woman's other household bargaining characteristics. This dataset includes information on three of these kinds of variables: whether a woman works outside the home, whether a woman has her name on the household papers, and whether a woman has access to a mobile phone in the household. There are a number of variables omitted from this household bargaining system due to unobserved characteristics not revealed in the data set. Due to the reverse causality and omitted variable bias of the OLS model, its coefficients will present biased estimates of the impact of women's mobile phone access on a household's consumption outcomes.

A valid Two Stage Least Squares design using instrumental variables could identify some heterogeneous source of variation in women's access to mobile phones that is not tied to some of the other household bargaining variables that are both *causes* and *effects* in the household bargaining system. This econometric model could more robustly determine how mobile phone access impacts consumption behaviors.

Common instruments in the literature on mobile phone access are the household's distance from their nearest mobile data network, cell phone tower, or other public re sources like rail or bus stations, public phones, or post offices (Abor, Amidu, and Issahaku 2018). A few papers use some distance variable with a variable for total number of mobile phones in the locality (Bair and Tritah 2019; Jack, Ray and Suri 2013). Tankari (2018) uses three instruments: the proportion of household heads who can read in any language, the presence of community radios, and the village distance from the administrative center, arguing that villages closer to the administrative center of their region sooner receive mobile phone service. The IHDS II Village Questionnaire contained characteristics about the village where the household is located. Three village-level instruments were found valid in these models. The first variable was distance to the headquarters of the district in which the village is located. Tankari (2018) finds that villages closer to the headquarters of the district receive mobile phone service and infrastructure sooner than villages farther from the headquarters. Then, this distance is relevant to women's mobile phone access but not necessarily related to household consumption decisions. The second instrument used is the distance of the household to the closest bus availability, whether that is a station, a bus stop, or a route along which someone can wave down the bus. Bus transportation and mobile phones both facilitate access to information and resources, meaning that mobile phones could be a substitute for bus transportation, an inferior good. This variable is not directly related to consumption decisions, so this can also be considered a valid instrument. The third instrument used is the number of years a village has had mobile phone access. This can identify the differences in levels of development between each village as well as most clearly relate to the rate of female mobile access. Since this is also not directly related to household consumption decisions, this third instrument is valid. This paper's regressions use these three instruments because the F statistic on most of the re gressions with these instruments was greater than 10, indicating a sufficiently strong joint instrument. These three instruments make up the first set of instruments included in the 2SLS regressions.

A second set of instruments is included in addition to the village-level instruments in order to more specifically identify *women's* mobile phone uptake. The village-level instruments are valid, but they can be criticized because they instrument for mobile phone uptake regardless of gender, and they do so at the village-level even though the data is at the household-level. Controlling for household ownership of a cell phone can isolate some part of male access to cell phones, but more instruments in addition to this control may work together for a stronger empirical strategy. I include three household-level bargaining power instruments: the economic status of a woman's natal family relative to her marital family, the difference in highest male and female education years, and a dummy for whether the household head is female. The household bargaining literature suggests these variables as good proxies for women's bargaining power that are not as affected by the other household bargaining variables as those are (mobile phone access, women's labor force participation rate, and whether her name is on the papers) as they are by these structural characteristics (Mohapatra and Simon 2017; Doss 2006 reviews empir ical literature). These characteristics are related to the household bargaining system, more likely as root causes of other household bargaining outcomes rather than effects. These characteristics are treated differently than the household bargaining controls discussed later on because these traits are usually set before a woman enters a household bargaining system. In other words, these are the closest ”exogenous” household bargaining variables available in the data set. Because these characteristics are more or less set by circumstances present be fore marriage, reverse causality with the consumption variables or the other bargaining variables is limited. These variables satisfy the relevance condition in that they are related to women's mobile phone uptake due to their presence in the household bargaining system.

The exclusion restriction merits further discussion. For these variables to be valid instruments, a convincing case must be made that they do not affect the consumption outcomes directly, only through their effect on women's mobile phone access. These ”systemic” variables (relative economic status of the woman's natal family, the difference in highest male and female education, and whether the household head is female), I argue, do not affect the consumption shares enough to make the instruments invalid. may affect a household's level of total consumption in that women indicating greater household bargaining power may have a greater impact on the consumption/savings tradeoff. Similarly, these variables may affect the allocation of consumption in the household according to the gender and consumption literature. Consumption shares of some necessities or goods that vary more on the day to day like energy, care goods, or household goods may not be affected by these consistent, persistent bargaining characteristics. Still, its potential effect on other consumption shares should not be understated. This paper argues that this effect is partly reduced by the inclusion of the valid village-level instruments. This issue still remains, however, and may overstate the results for the six-instrument models.

Below is the equation for the first stage equation regressing the dummy indicating at least one woman accessing a mobile phone on a vector of instruments and a vector of controls:

$$Mobile_i = \beta_0 + \beta_1 \boldsymbol{Instruments}_i + \beta_2 \boldsymbol{V}_i + \varepsilon_i$$

The second stage equation regresses the consumption outcome of interest on the estimates of the mobile access from the first stage and a vector of controls:

$$Consumption_i = \gamma_0 + \gamma_1 \widehat{Mobile}_i + \gamma_2 \boldsymbol{V}_i + \varepsilon_i$$

This paper clusters standard errors by primary sampling unit (PSU) and uses the weights provided in the Eligible Women questionnaire, a common strategy for the IHDS (Mohapatra and Simon 2017).

5 Results

This paper asks how a woman's access to a mobile phone in a household affects the allocation of that household's expenditures. This paper first reports results from Ordinary Least Squares regressions which can account for differences in factors that may also cause these changes in consumption shares, like wealth or village characteristics. This paper then accounts for the endogeneity between women's mobile phone access and household consumption by employing a 2SLS instrumental variables strategy. The results conclude with a robustness check considering how the role of mobile phones may differ in a household's bargaining system when the head is female.

5.1 Ordinary Least Squares Regressions

OLS regressions were first completed for all ten consumption share variables. Table 7 shows that five of the ten consumption shares had statistically significant differences between the two groups of households at least to the 15% level. The coefficients for the shares of finances and jewelry had positive coefficients while the shares for transportation, energy, and intoxicants/processed foods had negative coefficients. The magnitudes of these statistically significant coefficients range between 0.2 and 0.5 percentage points. The means of these shares are generally in the single digits as well, indicating that these coefficients may actually indicate an economically significant difference for the sampled households.

Table 8 shows the OLS regressions of logged total consumption or logged per capita consumption on the indicator of whether a household has any woman accessing a mobile phone. The coefficients on this indicator variable are not statistically significant, even at the 15% level. The OLS results presented here may not be the most accurate estimation of these variables' relationship. There could be some reverse causality between a woman's mobile phone access and the share of consumption in some cat-

For example, changing shares of energy, transportation, education, or other categories of consumption affect how much is left in the household budget to spend on greater mobile phone maintenance, more advanced functionalities, or more energy to keep them charged; these factors directly affect a woman's likelihood to access a mobile phone. Additionally, the household bargaining characteristics included in the regression may affect both the consumption outcomes and whether a woman accesses a mobile phone. An instrumental variable design can account for this endogeneity by identifying some source of heterogeneous variation in mobile phone uptake.

Table 7: OLS Consumption Shares - All Controls

	(1) Food	(2) Education	(3) Finances	(4) Transportation	(5) Energy	(6) Household Goods	(7) Jewelry	(8) Discretionary	(9) Care Goods	(10) Intoxicants and Processed Foods
Any Woman with Mobile Access	-0.00331	0.00308	0.00438***	-0.00377*	-0.00425**	-0.0000155	0.00224^{+}	-0.000353	0.00429	-0.00229***
	(-0.88)	(1.39)	(3.84)	(-1.87)	(-2.57)	(-0.01)	(1.62)	(-0.14)	(1.40)	(-2.71)
Household Owns Cell Phone	-0.0100*	-0.000427	-0.00115	-0.00191	0.00700***	0.00183	0.000952	0.00511^{+}	0.000794	-0.00219^{+}
	(-1.95)	(-0.18)	(-0.95)	(-1.04)	(3.01)	(1.20)	(0.84)	(1.55)	(0.16)	(-1.63)
Total Household Assets (Index 0-33)	-0.00848***	0.00272***	0.00161***	0.00457***	0.000264	0.00146***	0.000658***	0.000661***	-0.00258***	-0.000885***
	(-19.10)	(10.13)	(9.69)	(16.64)	(1.34)	(7.83)	(4.71)	(2.59)	(-6.36)	(-7.95)
Highest Adult Education in a Household	-0.00177***	0.00143***	0.000591***	0.000542**	-0.0000338	-0.000184	-0.000168	-0.000553*	0.000581^{+}	-0.000437***
	(-4.43)	(4.80)	(4.55)	(2.45)	(-0.20)	(-1.13)	(-1.24)	(-1.90)	(1.56)	(-4.63)
Household Ratio of Adults:Children	0.00546*	0.00614***	0.00400***	0.00171	-0.00206*	-0.00291***	-0.00307***	-0.00984***	0.000749	-0.000166
	(1.95)	(4.26)	(4.45)	(1.14)	(-1.86)	(-3.62)	(-4.69)	(-4.26)	(0.28)	(-0.28)
Size of Household	0.00452***	0.000674^{+}	-0.00199***	-0.00128***	-0.00326***	-0.0000759	-0.000305	-0.000588	0.00203***	0.000288**
	(5.74)	(1.56)	(-6.49)	(-2.96)	(-10.55)	(-0.28)	(-1.18)	(-1.12)	(3.02)	(2.19)
Log of Total HH Income	-0.00555***	-0.00143	0.00452***	0.00656***	-0.00389***	0.000822	0.00315***	0.00508***	-0.0101***	0.000794^{+}
	(-2.61)	(-1.08)	(6.02)	(6.00)	(-4.51)	(0.96)	(3.30)	(3.89)	(-5.46)	(1.58)
Village: Food for Work Program	-0.0108**	-0.00416*	0.00195	0.00384	0.00261	0.00364**	0.000798	0.00705**	-0.00557	0.000590
	(-2.00)	(-1.87)	(1.10)	(1.34)	(1.15)	(2.05)	(0.54)	(2.00)	(-1.41)	(0.52)
Village: MG-NREGA	-0.000167	0.00740^{+}	0.00275	-0.00731	-0.0129**	0.00546**	-0.0000902	0.00588	-0.00294	0.00191
	(-0.01)	(1.57)	(0.85)	(-1.02)	(-2.28)	(2.14)	(-0.02)	(1.11)	(-0.27)	(1.08)
Village: Other Govt Employment Guarantee	0.00261	-0.00134	0.00160	-0.000879	0.00174	0.00139	0.000586	0.00597*	-0.00963**	-0.00204*
	(0.49)	(-0.63)	(0.96)	(-0.38)	(0.61)	(0.81)	(0.34)	(1.78)	(-2.40)	(-1.88)
Village: Mahila Mandal (Women's Assoc.)	-0.00888*	0.000106	0.00120	-0.000933	0.00812***	0.00100	-0.00274**	0.000715	0.000569	0.000833
	(-1.83)	(0.05)	(0.87)	(-0.45)	(3.40)	(0.63)	(-2.12)	(0.25)	(0.15)	(0.81)
Electricity Access Hours/Day	0.000255	-0.000574***	-0.0000346	-0.000138	0.000534***	0.000130	-0.0000664	0.000495*	-0.000926***	0.000325***
	(0.72)	(-3.50)	(-0.30)	(-0.87)	(2.73)	(0.97)	(-0.69)	(1.83)	(-3.29)	(4.15)
Any Female Name on Household Papers	0.00506	-0.0000304	0.00158	-0.00151	-0.00379*	0.00376^{+}	0.00271^{+}	-0.00760**	-0.0000250	-0.000147
	(1.00)	(-0.01)	(1.01)	(-0.64)	(-1.70)	(1.48)	(1.50)	(-2.32)	(-0.01)	(-0.13)
Any Woman Works Outside Home	-0.00220	-0.00147	-0.00193	-0.00473**	0.00246	0.00158	-0.00399**	0.00372	0.00628^{+}	0.000291
	(-0.46)	(-0.61)	(-1.32)	(-2.46)	(1.32)	(0.98)	(-2.55)	(0.69)	(1.55)	(0.29)
Constant	0.633***	0.00454	-0.0578***	-0.0594***	0.161***	-0.000672	-0.0289***	-0.00782	0.325***	0.0311***
	(24.62)	(0.32)	(-6.78)	(-4.52)	(14.96)	(-0.08)	(-2.84)	(-0.57)	(14.83)	(5.89)
Observations	16447	16447	16447	16447	16447	16447	16447	16447	16447	16447

t statistics in parentheses
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include controls for region.

5.2 2SLS - Logged Total and Per Capita Consumption

Table 9 presents 2SLS regressions of any woman's mobile phone access on logged total consumption and logged per capita consumption. These use the three village-level instrumental variables, weighted data, and clustered standard errors by Primary Sampling Unit (PSU). When incorporating the full panel of controls, F statistics are much larger than 10 but no coefficients on the dummy for women's mobile access are statistically significant. It is possible that these instruments are valid for consumption allocations but not total consumption measures. These village-level characteristics have a clear relationship to mobile phone uptake, albeit uptake not specific to women, but these may be related to total con sumption in other ways. Some controls account for village-level development separately, but the residual correlation between these instruments and development factors that affect total consumption may make this specification statistically weak.

Table 8: OLS Logged Consumption - All Controls

	(1) Total Consumption	(2) Per Capita Consumption
Any Woman with Mobile Access	-0.0133	-0.00614
	(-1.00)	(-0.46)
Household Owns Cell Phone	0.101***	0.0589***
	(5.81)	(3.24)
Total Household Assets (Index 0-33)	0.0490***	0.0508***
	(29.71)	(29.15)
Highest Adult Education in a Household	0.00509***	0.0000983
	(3.55)	(0.07)
Household Ratio of Adults:Children	-0.0454***	-0.108***
	(-4.69)	(-10.12)
Size of Household	0.0745***	-0.0942***
	(26.73)	(-29.09)
Log of Total HH Income	0.112***	0.103***
	(13.55)	(12.12)
Village: Food for Work Program	0.0210	0.0152
	(0.94)	(0.68)
Village: MG-NREGA	-0.0158	-0.0294
	(-0.34)	(-0.58)
Village: Other Govt Employment Guarantee	-0.0446**	-0.0434*
	(-2.04)	(-1.96)
Village: Mahila Mandal (Women's Assoc.)	-0.00737	-0.0138
	(-0.39)	(-0.72)
Electricity Access Hours/Day	-0.00580***	-0.00635***
	(-4.00)	(-4.33)
Any Female Name on Household Papers	-0.0293^{+}	-0.0134
	(-1.62)	(-0.74)
Any Woman Works Outside Home	-0.0424**	-0.0498***
	(-2.57)	(-2.94)
Constant	9.050***	8.576***
	(93.58)	(83.92)
Observations	16447	16447

t statistics in parentheses
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include controls for region.

Table 10 presents the 2SLS regressions of logged total consumption and logged consumption per capita with the six instrumental variables. These models show negative coefficients on women's mobile phone access that are statistically significant to the 1% level. Model 1 suggests that when all controls are included, logged total consumption is lower for households with at least one woman accessing a mobile phone by 0.327. Converting the logged value to a percentage, it seems that households with a woman accessing a mobile phone have decreased total consumption by 27.9%. Similarly, per capita consumption is lower for these households by 28.3%. The decreases present in both models suggest that there could be increased saving of income for households with a woman accessing a mobile phone. It is plausible that when a woman increases her bargaining power through access to a mobile phone, she advocates for more saving rather than spending of household income. The literature on different savings preferences by gender is mixed with very little focusing on rural India, so this reasoning is purely speculative. Discussion of the differences in consumption shares may be able to contextualize these differences.

5.3 2SLS - Consumption Shares

Research on mobile phones in the developing world has found a wide range of benefits. Much recent literature focuses on the impacts of mobile phones or mobile money on women's autonomy indicators or a range of health outcomes (Rotondi et al. 2020; Mo han et al. 2020; Kruse et al. 2019; Biswas et al. 2021; Stark 2020). Some have found that mobile phones lower information costs for firms, increase cooperation, and decrease price dispersion in local markets (Jensen 2007; Aker and Mbiti 2010; Goyal 2008). The social networks enabled by mobile phones can help disperse information, opportunity, and assets for households. The reduction in travel and information costs has been one of the most significant benefits of mobile phone use for households, individuals, and firms alike. Time poverty, or the limited time one has for economically productive activities, plagues women in the developing world (Hyde, Greene, and Darmstadt 2020). Mobile phones can share information and resources with women and households that would otherwise take much time, money, and distance to acquire (Muto and Yamano 2009; Aker and Mbiti 2010; Jensen 2007; Overa 2006). For example, Goodman 2005 found that mobile phones enabled households to maintain social networks that, in turn, shared information and economic opportunity. Samuel et al. (2005) similarly found that the types of information spread to households were business information, friends and family contacts, and job prospects.

A strand of literature connects access to mobile phones in the developing world to household economic welfare. Adoption of mobile phones and mobile money has been found to increase the level and growth rate of household per capita consumption, level of overall household expenditures, and household real consumption while lowering poverty incidence across many developing countries (Munyegera and Matsumoto 2016; Labonne and Chase 2009; Tankari 2018; Beuermann, McKelvey, and Vakis 2012). Other papers have found that mobile phones enable risk sharing across social networks, increased financial resiliency, and long-term household consumption increases with greater effects for female-headed households (Jack and Suri 2014; Rettie, 2008; Goodman 2005; Duncombe 2014; Riley 2018; Suri and Jack 2016; Sekabira and Qaim 2017). Many previous studies did not disaggregate their findings by gender, however. Some that do suggest that while per capita household consumption increased with mobile phone ownership, there was not a statistically significant difference in effects for male versus female mobile phone owners (Tankari 2018).

Tables 11-14 show the results of the 2SLS regressions on each category of consumption, starting with no control variables and adding them in groups. Regional controls are included in all of these regressions because of their high statistical significance in all iterations of the regressions. Table 11 presents the 10 models without other additional controls. Consumption shares of education, finances, transportation, and vices are statistically significant while most of the F-statistics are above 10. These factors suggest that there is some baseline variance in consumption allocations between the groups even when using 2SLS and that more controls should be added to examine it more precisely. Table 12 adds Group 1 control variables describing the household's composition. The Group 2 controls added in Table 13 are village characteristics, and the remaining controls added in Table 14 are household bargaining characteristics.

Table 11: 2SLS Consumption Shares - Village Instruments, No Control

	(1) Food	(2) Education	(3) Finances	(4) Transportation	(5) Energy	(6) Household Goods	(7) Jewelry	(8) Discretionary	(9) Care Goods	(10) Intoxicants and Processed Foods
Any Woman with Mobile Access	-0.0466	0.0598***	0.0714***	0.0531**	-0.0236	0.0179	-0.00663	-0.0192	-0.0747+	-0.0314***
	(-1.00)	(2.84)	(3.85)	(2.34)	(-0.83)	(1.30)	(-0.52)	(-0.72)	(-1.60)	(-2.61)
Constant	0.496***	0.0122	-0.0248***	0.0266**	0.113***	0.0234***	0.0120*	0.0682***	0.227***	0.0460***
	(20.23)	(1.11)	(-2.61)	(2.24)	(7.31)	(3.22)	(1.70)	(5.10)	(8.84)	(7.28)
Observations	16447	16447	16447	16447	16447	16447	16447	16447	16447	16447
F	30.44	18.45	16.95	14.14	14.86	6.040	3.915	13.13	21.95	4.423

t statistics in parentheses
+ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include region controls.

Table 10: 2SLS Logged Consumption - All Instruments, All Controls

	(1) Per Capita Consumption	(2) Total Consumption
Any Woman with Mobile Access	-0.261	-0.295
	(-0.70)	(-0.81)
Household Owns Cell Phone	0.162	0.215^{+}
	(1.09)	(1.49)
Household Ratio of Adults:Children	-0.0928***	-0.0284
	(-3.73)	(-1.17)
Size of Household	-0.0977***	0.0707***
	(-14.94)	(11.62)
Log of Total HH Income	0.106***	0.115***
	(11.07)	(12.37)
Total Household Assets (Index 0-33)	0.0543***	0.0529***
	(10.45)	(10.60)
Highest Adult Education in a Household	0.00256	0.00782**
	(0.62)	(1.96)
Village: Food for Work Program	0.00437	0.00897
	(0.15)	(0.31)
Village: MG-NREGA	-0.00510	0.0111
	(-0.09)	(0.19)
Village: Other Govt Employment Guarantee	-0.0503**	-0.0523**
	(-2.13)	(-2.25)
Village: Mahila Mandal (Women's Assoc.)	-0.0143	-0.00788
	(-0.72)	(-0.40)
Electricity Access Hours/Day	-0.00493*	-0.00423^{+}
	(-1.84)	(-1.62)
Any Woman Works Outside Home	-0.0353	-0.0264
	(-1.32)	(-1.03)
Any Female Name on Household Papers	-0.0104	-0.0259
	(-0.54)	(-1.37)
Constant	8.509***	8.976***
	(62.32)	(66.33)
Observations	16447	16447
F	183.5	271.3

31

t statistics in parentheses
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: 2SLS Consumption Shares - Village Instruments, Group 1 Control

	(1) Food	(2) Education	(3) Finances	(4) Transportation	(5) Energy	(6) Household Goods	(7) Jewelry	(8) Discretionary	(9) Care Goods	(10) Intoxicants and Processed Foods
Any Woman with Mobile Access	0.329**	-0.0107	0.0447^{+}	-0.103^{+}	-0.0243	-0.0177	-0.0567^{+}	-0.0896^{+}	-0.0615	-0.0101
	(2.00)	(-0.33)	(1.51)	(-1.59)	(-0.47)	(-0.73)	(-1.51)	(-1.49)	(-0.73)	(-0.51)
Total Household Assets (Index 0-33)	-0.0172***	0.00297***	0.000575	0.00710***	0.00103	0.00200***	0.00227**	0.00310**	-0.00115	-0.000698
	(-4.16)	(3.57)	(0.75)	(4.31)	(0.79)	(3.21)	(2.33)	(2.02)	(-0.55)	(-1.34)
Highest Adult Education in a Household	-0.00547***	0.00164***	0.000163	0.00166**	0.000174	-0.00000580	0.000518	0.000390	0.00132	-0.000396^{+}
	(-2.75)	(3.38)	(0.44)	(1.96)	(0.28)	(-0.02)	(1.09)	(0.54)	(1.27)	(-1.64)
Household Ratio of Adults:Children	-0.0118	0.00702***	0.00184	0.00708*	-0.00137	-0.00217	0.0000654	-0.00521	0.00440	0.000198
	(-1.17)	(3.07)	(0.95)	(1.70)	(-0.50)	(-1.38)	(0.03)	(-1.24)	(0.89)	(0.16)
Size of Household	0.00679***	0.000577	-0.00169***	-0.00205***	-0.00327***	-0.000190	-0.000765^{+}	-0.00116^{+}	0.00157^{+}	0.000185
	(3.75)	(1.16)	(-4.55)	(-3.01)	(-5.83)	(-0.57)	(-1.61)	(-1.56)	(1.57)	(0.91)
Log of Total HH Income	-0.00875**	-0.00119	0.00396***	0.00746***	-0.00390***	0.000866	0.00361***	0.00584***	-0.00869***	0.000796^{+}
	(-2.46)	(-0.80)	(4.84)	(5.32)	(-3.60)	(0.93)	(3.10)	(3.78)	(-4.07)	(1.53)
Constant	0.621***	0.00385	-0.0543***	-0.0682***	0.162***	0.0106	-0.0290***	0.0106	0.307***	0.0365***
	(19.09)	(0.28)	(-6.57)	(-5.27)	(16.30)	(1.20)	(-2.87)	(0.75)	(16.03)	(7.29)
Observations	16447	16447	16447	16447	16447	16447	16447	16447	16447	16447
F	50.89	39.26	33.99	76.99	23.12	12.88	8.446	9.472	22.01	22.62

t statistics in parentheses
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include region controls.

Table 13: 2SLS Consumption Shares - Village Instruments, Groups 1 and 2 Control

	(1) Food	(2) Education	(3) Finances	(4) Transportation	(5) Energy	(6) Household Goods	(7) Jewelry	(8) Discretionary	(9) Care Goods	(10) Intoxicants and Processed Foods
Any Woman with Mobile Access	0.342**	-0.0104	0.0444^{+}	-0.0990^{+}	-0.0365	-0.0249	-0.0553^{+}	-0.104^{+}	-0.0414	-0.0140
	(1.97)	(-0.33)	(1.47)	(-1.55)	(-0.70)	(-0.96)	(-1.44)	(-1.62)	(-0.51)	(-0.69)
Total Household Assets (Index 0-33)	-0.0174***	0.00307***	0.000608	0.00697***	0.00122	0.00214***	0.00222**	0.00331**	-0.00150	-0.000662
	(-4.09)	(3.74)	(0.80)	(4.37)	(0.95)	(3.31)	(2.28)	(2.06)	(-0.77)	(-1.28)
Highest Adult Education in a Household	-0.00571***	0.00159***	0.000146	0.00165*	0.000351	0.0000921	0.000520	0.000627	0.00105	-0.000318
	(-2.66)	(3.38)	(0.38)	(1.93)	(0.55)	(0.27)	(1.04)	(0.79)	(1.02)	(-1.26)
Household Ratio of Adults:Children	-0.0129	0.00688***	0.00185	0.00690*	-0.000365	-0.00173	-0.0000481	-0.00421	0.00312	0.000495
	(-1.21)	(2.96)	(0.94)	(1.66)	(-0.13)	(-1.03)	(-0.02)	(-0.93)	(0.64)	(0.39)
Size of Household	0.00685***	0.000569	-0.00172***	-0.00200***	-0.00337***	-0.000248	-0.000723^{+}	-0.00123*	0.00171*	0.000169
	(3.78)	(1.21)	(-4.73)	(-3.07)	(-6.35)	(-0.73)	(-1.57)	(-1.65)	(1.82)	(0.85)
Log of Total HH Income	-0.00934**	-0.00134	0.00401***	0.00741***	-0.00342***	0.00109	0.00357***	0.00642***	-0.00933***	0.000933*
	(-2.53)	(-0.94)	(4.87)	(5.28)	(-3.13)	(1.17)	(3.02)	(3.98)	(-4.41)	(1.77)
Village: Food for Work Program	0.00644	-0.00480*	0.00395	-0.000743	0.000985	0.00223	-0.00210	0.00198	-0.00797	0.0000379
	(0.46)	(-1.82)	(1.36)	(-0.13)	(0.29)	(0.99)	(-0.71)	(0.38)	(-1.29)	(0.03)
Village: MG-NREGA	-0.0349	0.00866^{+}	-0.00137	0.00193	-0.00953	0.00813**	0.00549	0.0165^{+}	0.00205	0.00309
	(-1.14)	(1.46)	(-0.27)	(0.20)	(-1.29)	(2.14)	(0.89)	(1.60)	(0.14)	(1.12)
Village: Other Govt Employment Guarantee	0.0122	-0.00170	0.00275	-0.00348	0.000771	0.000728	-0.000925	0.00294	-0.0110**	-0.00237*
	(1.15)	(-0.75)	(1.33)	(-0.96)	(0.24)	(0.37)	(-0.41)	(0.65)	(-2.36)	(-1.81)
Village: Mahila Mandal (Women's Assoc.)	-0.00798	0.000113	0.00134	-0.00104	0.00803***	0.000843	-0.00281^{+}	0.000442	0.000264	0.000789
	(-0.98)	(0.05)	(0.83)	(-0.38)	(3.30)	(0.50)	(-1.56)	(0.13)	(0.07)	(0.72)
Electricity Access Hours/Day	-0.00113	-0.000519**	-0.000201	0.000256	0.000652**	0.000219	0.000154	0.000924**	-0.000733^{+}	0.000383***
	(-1.21)	(-2.42)	(-1.20)	(0.84)	(2.12)	(1.26)	(0.77)	(2.48)	(-1.52)	(3.12)
Constant	0.671***	0.00288	-0.0530***	-0.0714***	0.158***	-0.00199	-0.0343***	-0.0204	0.320***	0.0289***
	(14.52)	(0.19)	(-5.36)	(-4.37)	(12.71)	(-0.21)	(-2.67)	(-1.16)	(13.05)	(5.04)
Observations	16447	16447	16447	16447	16447	16447	16447	16447	16447	16447
F	33.79	27.91	22.97	53.27	18.72	9.952	6.135	7.140	16.65	15.68

t statistics in parentheses
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include region controls.

Table 14: 2SLS Consumption Shares - Village Instruments, All Control

	(1) Food	(2) Education	(3) Finances	(4) Transportation	(5) Energy	(6) Household Goods	(7) Jewelry	(8) Discretionary	(9) Care Goods	(10) Intoxicants and Processed Foods
Any Woman with Mobile Access	0.311**	-0.00648	0.0405^{+}	-0.0903^{+}	-0.0386	-0.0188	-0.0492^{+}	-0.0829^{+}	-0.0551	-0.0104
	(2.13)	(-0.22)	(1.51)	(-1.64)	(-0.78)	(-0.83)	(-1.46)	(-1.52)	(-0.69)	(-0.56)
Household Owns Cell Phone	-0.137**	0.00344	-0.0158^{+}	0.0331^{+}	0.0209	0.00942	0.0218^{+}	0.0385*	0.0248	0.00111
	(-2.33)	(0.28)	(-1.45)	(1.50)	(1.05)	(1.02)	(1.60)	(1.74)	(0.77)	(0.14)
Total Household Assets (Index 0-33)	-0.0128***	0.00285***	0.00111***	0.00577***	0.000737	0.00172***	0.00137***	0.00180**	-0.00176^{+}	-0.000773***
	(-6.39)	(6.01)	(2.76)	(7.21)	(1.11)	(4.97)	(2.77)	(2.32)	(-1.64)	(-2.75)
Highest Adult Education in a Household	-0.00482***	0.00152***	0.000240	0.00138**	0.000299	-0.00000162	0.000330	0.000247	0.00116	-0.000358*
	(-3.05)	(3.66)	(0.77)	(2.06)	(0.56)	(-0.01)	(0.84)	(0.42)	(1.28)	(-1.71)
Household Ratio of Adults:Children	-0.0136	0.00671***	0.00180	0.00695*	0.0000265	-0.00178	0.0000520	-0.00484	0.00435	0.000327
	(-1.36)	(2.81)	(0.92)	(1.73)	(0.01)	(-1.08)	(0.02)	(-1.09)	(0.81)	(0.25)
Size of Household	0.00877***	0.000544	-0.00150***	-0.00246***	-0.00373***	-0.000330	-0.00100*	-0.00171*	0.00122	0.000177
	(3.83)	(0.96)	(-3.45)	(-2.91)	(-4.76)	(-0.80)	(-1.67)	(-1.91)	(0.93)	(0.63)
Log of Total HH Income	-0.00840**	-0.00134	0.00419***	0.00734***	-0.00358***	0.000992	0.00362***	0.00583***	-0.00952***	0.000868*
	(-2.55)	(-0.99)	(5.27)	(5.61)	(-3.42)	(1.12)	(3.19)	(4.17)	(-4.52)	(1.69)
Village: Food for Work Program	0.00263	-0.00457*	0.00349	0.000154	0.00115	0.00285	-0.00139	0.00354	-0.00810	0.000243
	(0.23)	(-1.85)	(1.34)	(0.03)	(0.37)	(1.39)	(-0.55)	(0.81)	(-1.38)	(0.18)
Village: MG-NREGA	-0.0301	0.00831^{+}	-0.000694	0.000939	-0.00962	0.00725**	0.00481	0.0138^{+}	0.00271	0.00268
	(-1.12)	(1.48)	(-0.15)	(0.10)	(-1.37)	(2.11)	(0.86)	(1.57)	(0.19)	(1.06)
Village: Other Govt Employment Guarantee	0.0112	-0.00160	0.00258	-0.00323	0.000810	0.000878	-0.000812	0.00372	-0.0112**	-0.00226*
	(1.18)	(-0.71)	(1.29)	(-0.96)	(0.26)	(0.47)	(-0.38)	(0.93)	(-2.40)	(-1.79)
Village: Mahila Mandal (Women's Assoc.)	-0.00831	0.0000890	0.00127	-0.00109	0.00806***	0.000970	-0.00283*	0.000565	0.000461	0.000818
	(-1.13)	(0.04)	(0.81)	(-0.42)	(3.32)	(0.59)	(-1.68)	(0.18)	(0.12)	(0.77)
Electricity Access Hours/Day	-0.00150^{+}	-0.000521**	-0.000237	0.000345	0.000726**	0.000235	0.000221	0.000956**	-0.000595	0.000371***
	(-1.55)	(-2.18)	(-1.30)	(1.02)	(2.01)	(1.29)	(1.04)	(2.55)	(-1.03)	(2.68)
Any Female Name on Household Papers	0.00129	0.0000843	0.00115	-0.000471	-0.00338^{+}	0.00398^{+}	0.00332^{+}	-0.00661*	0.000688	-0.0000493
	(0.17)	(0.03)	(0.68)	(-0.16)	(-1.45)	(1.55)	(1.59)	(-1.79)	(0.16)	(-0.04)
Any Woman Works Outside Home	-0.0201*	-0.000929	-0.00398*	0.000184	0.00441	0.00265	-0.00107	0.00841	0.00966*	0.000755
	(-1.89)	(-0.32)	(-1.92)	(0.04)	(1.39)	(1.23)	(-0.42)	(1.42)	(1.65)	(0.49)
Constant	0.716***	0.00202	-0.0483***	-0.0822***	0.152***	-0.00562	-0.0424***	-0.0296^{+}	0.310***	0.0290***
	(13.19)	(0.12)	(-4.37)	(-4.20)	(9.18)	(-0.53)	(-2.73)	(-1.49)	(9.93)	(4.23)
Observations	16447	16447	16447	16447	16447	16447	16447	16447	16447	16447
F	35.24	25.43	22.11	49.48	16.72	10.93	7.229	8.410	13.87	13.44

t statistics in parentheses
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions include controls for region.

Table 14 shows these consumption share regressions with the three village-level instruments and all controls. These regressions suggest that the food and finance shares are higher for households with women's mobile phone access, but the shares of transportation, jewelry, and discretionary spending are lower. First, Column 1 suggests that households with at least one woman accessing a mobile phone have higher food shares on average by 33.7 percentage points, statistically significant to the 5% level. This is a very high magnitude relative to the mean food share in the sample. The mean food share was 46.5%, so this regression result would suggest that households with one woman accessing a mobile phone have a roughly 72% higher consumption share of food. This seems unlikely given the relatively low level of annual consumption of these rural households: the average total consumption of |106,383 (2011) for a household in the sample, converting to roughly USD 2,638 (2022). If annual consumption is so low to begin with, it seems suspect that households with a woman accessing a mobile phone would have so much available income to reallocate towards food. This partly motivates the search for a more statistically precise econometric method, including the six instruments discussed later on.

The other shares were only statistically significant to the 15% level, but their coefficients are still interesting to discuss. Column 3 suggests that households with at least one woman accessing a mobile phone spend five percentage points more on household finances than households without. This could indicate an increase in general household financial activity such as spending on consumer taxes, household loan payments, or various rents. The transportation share was found to be almost 10 percentage points lower for households with at least one woman accessing a mobile phone. This could signal how mobile phones substitute for transportation by lowering transaction costs for gathering information, thus leading households to spend less on unnecessary, risky transportation. The discretionary spending share was found to be 8.29 percentage points lower for households with women's mobile phone access. The instruments weakly identify the discretionary spending model according to an F-statistic of 8.410, but the coefficient is still interesting for discussion. This could indicate that when women have more bargaining power, former discretionary spending decreases and is saved.

Lastly, Column 7 of Table 14 shows that the jewelry share was lower by 4.92 percentage points for households with women's mobile phone access. The instruments weakly identify the jewelry model according to an F-statistic of 7.23, but the coefficient is still interesting to discuss as well.

Jewelry is often used as a social status symbol in rural India, so mobile phones could have two negative effects on this consumption share. Some literature suggests that mobile phones are markers of social status in rural communities, particularly when mobile phones are shared in the community (Hahn and Kibora 2008). It is also possible that mobile phones introduce women to attitudes and expectations from other areas, making them less interested in spending on jewelry. Further research with more precise instruments can reveal more about these shifting consumption patterns. Overall, the decreases in the transportation, discretionary, and jewelry consumption shares could reveal why the total and per capita consumption measures decreased according to the 2SLS regressions in Table 10.

Table 15: 2SLS Consumption Shares - All Instruments, All Control

	(1) Food	(2) Education	(3) Finances	(4) Transportation	(5) Energy	(6) Household Goods	(7) Jewelry	(8) Discretionary	(9) Care Goods	(10) Intoxicants and Processed Foods
Any Woman with Mobile Access	0.0462**	-0.00675	0.0156**	-0.0200*	0.00861	-0.00306	-0.00731	-0.0305	-0.00263	-0.0000893
	(2.13)	(-0.50)	(2.06)	(-1.76)	(0.88)	(-0.40)	(-1.06)	(-1.39)	(-0.14)	(-0.02)
Household Owns Cell Phone	-0.0292***	0.00220	-0.00574*	0.00398	0.00182	0.00234	0.00429^{+}	0.0168**	0.00717	-0.00372^{+}
	(-3.33)	(0.40)	(-1.84)	(0.83)	(0.46)	(0.69)	(1.54)	(2.11)	(0.88)	(-1.56)
Total Household Assets (Index 0-33)	-0.00895***	0.00279***	0.00146***	0.00494***	0.000147	0.00146***	0.000773***	0.000909***	-0.00272***	-0.000819***
	(-18.54)	(8.56)	(7.48)	(16.19)	(0.73)	(7.39)	(4.55)	(2.66)	(-6.15)	(-6.03)
Highest Adult Education in a Household	-0.00257***	0.00174***	0.000494***	0.000682**	-0.000206	-0.000146	-0.000111	-0.000103	0.000831*	-0.000614***
	(-5.28)	(4.61)	(3.11)	(2.25)	(-0.96)	(-0.82)	(-0.71)	(-0.27)	(1.93)	(-5.28)
Household Ratio of Adults:Children	0.00336	0.00468***	0.00448***	0.00423**	-0.00232*	-0.00305***	-0.00391***	-0.0108***	0.00137	0.00199**
	(0.95)	(2.85)	(4.08)	(2.18)	(-1.83)	(-2.68)	(-4.05)	(-3.06)	(0.44)	(2.53)
Size of Household	0.00494***	0.000842*	-0.00190***	-0.00172***	-0.00301***	-0.000144	-0.000217	-0.000710^{+}	0.00218***	-0.000268*
	(5.89)	(1.82)	(-5.74)	(-3.54)	(-8.64)	(-0.49)	(-0.84)	(-1.46)	(2.99)	(-1.89)
Log of Total HH Income	-0.00708***	-0.00139	0.00451***	0.00669***	-0.00398***	0.000850	0.00311***	0.00523***	-0.00845***	0.000518
	(-3.27)	(-0.96)	(5.82)	(5.87)	(-4.47)	(0.97)	(3.30)	(4.18)	(-4.35)	(1.01)
Village: Food for Work Program	-0.00912^{+}	-0.00477*	0.00261	0.00291	0.00257	0.00300*	0.000857	0.00681*	-0.00554	0.000678
	(-1.61)	(-1.94)	(1.37)	(0.91)	(1.05)	(1.67)	(0.53)	(1.93)	(-1.34)	(0.57)
Village: MG-NREGA	-0.00453	0.00849*	0.00127	-0.00530	-0.0137**	0.00636**	0.000102	0.00813	-0.00276	0.00190
	(-0.31)	(1.69)	(0.38)	(-0.71)	(-2.51)	(2.47)	(0.03)	(1.34)	(-0.24)	(1.00)
Village: Other Govt Employment Guarantee	0.00434	-0.00264	0.00170	-0.000970	0.00305	0.000964	-0.000288	0.00451	-0.00887**	-0.00179^{+}
	(0.77)	(-1.19)	(0.99)	(-0.39)	(1.05)	(0.55)	(-0.18)	(1.35)	(-2.08)	(-1.63)
Village: Mahila Mandal (Women's Assoc.)	-0.00971*	0.000515	0.00140	-0.00103	0.00797***	0.00113	-0.00183	-0.000393	0.00110	0.000857
	(-1.95)	(0.23)	(0.99)	(-0.48)	(3.25)	(0.69)	(-1.38)	(-0.13)	(0.29)	(0.82)
Electricity Access Hours/Day	-0.0000862	-0.000520***	-0.000129	-0.0000320	0.000396*	0.000159	0.0000375	0.000678***	-0.000810***	0.000307***
	(-0.23)	(-3.13)	(-1.04)	(-0.18)	(1.81)	(1.18)	(0.34)	(2.86)	(-2.66)	(3.66)
Any Female Name on Household Papers	0.00447	-0.00140	0.000934	-0.00142	-0.00499**	0.00299	0.00312*	-0.00780**	0.00327	0.000828
	(0.85)	(-0.44)	(0.59)	(-0.55)	(-2.25)	(1.12)	(1.67)	(-2.21)	(0.71)	(0.66)
Any Woman Works Outside Home	-0.00451	-0.000857	-0.00214	-0.00383*	0.00125	0.00191	-0.00333**	0.00546	0.00436	0.00169^{+}
	(-0.93)	(-0.32)	(-1.31)	(-1.82)	(0.66)	(1.11)	(-2.53)	(1.04)	(1.08)	(1.50)
Constant	0.658***	0.00190	-0.0555***	-0.0647***	0.163***	-0.000385	-0.0302***	-0.0124	0.302***	0.0377***
	(24.14)	(0.12)	(-6.22)	(-4.62)	(14.92)	(-0.04)	(-2.82)	(-0.86)	(12.84)	(6.91)
Observations	15477	15477	15477	15477	15477	15477	15477	15477	15477	15477
F	72.73	23.76	23.09	59.36	14.42	9.770	8.392	8.312	12.57	17.74

t statistics in parentheses

$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All regressions include controls for region.

Table 15 shows the ten consumption share regressions with the three village-level instruments and the three bargaining-power instruments. Compared to the three-instrument model, these results are generally smaller in magnitude but more statistically significant. The coefficient for the food share remained positive and became statistically significant to the 5% level. Column 1 suggests that households with a woman accessing a mobile phone have a food share higher by 4.62 percentage points. Compared to the mean of the food share for the sample, this coefficient implies just a 9.94% higher food share for households with women's mobile phone access. Next, the jewelry and discretionary spending shares were no longer statistically significant to the 15% level, but the coefficients for the finances and transportation shares became statistically sig nificant to the 5% and 10 % level, respectively. Column 3 suggests that households with women's mobile phone access have higher consumption shares of household finance by 1.56 percentage points. Relative to the sample mean finance share of 2.0%, this suggests roughly a 78% increase for these households. Lastly, Column 4 shows that households with a woman accessing a mobile phone had smaller transportation shares by 2.00 percentage points, significant to the 10% level. Relative to the sample mean of the transportation share, 6.28%, this constitutes a 32% smaller consumption share for these households. This could still be a practically significant difference, implying that mobile phones could be substituting for transportation.

One thing to note about understanding these shifts in consumption shares is considering whether the shares have changed because the amount allocated to the consumption category has changed or if the amount allocated to other consumption categories has changed enough to make that share seem to grow or shrink relatively. In other words, it is unclear whether the sizes of the slices of the pie have changed or whether the overall size of the pie has changed, thus affecting the size of each slice. The total consumption regressions suggest that the total amount of consumption was 28% lower for households with women's mobile phone access. It is unlikely that the consumption levels of all shares decreased by the same proportion in order to achieve this result. Indeed, many of the coefficients on the consumption share regressions in Table 15 are negative although they are statistically insignificant. Perhaps small decreases across all of these categories led to an overall decrease in consumption of 28% on average. Different data and more precise instruments may explain this in further research.

5.4 *Female Household Head Robustness Check*

The bargaining dynamics between a male household head and his wife are undoubtedly different from those between a female household head and the senior members of her household. It can be argued that female household heads have more bargaining power than the senior females in the households with a male head, indicating that female heads' preferences are likely realized more often than those of the senior female. Next, consider the effects on mobile phones on consumption decisions directly and through its effects on bargaining power. It has been argued in this paper that mobile phones can affect women's access to information, freedom of movement, and connections to social networks. It is helpful to refer back to the collective household bargaining model:

$$U_1(C_1,Q) + Z \times U_2(C_2,Q)$$

$$U_1(C_1,Q) + Z \times U_2(C_2,Q)$$

Those channels directly affect consumption preferences and self-confidence to change women's utility function, but they may also combine in such a way as to increase their bargaining power relative to a male household head, affecting their relative weight Z. If mobile phones are a way for women in male-headed households to gain bargaining power *and* realize some direct effects on preferences and activity, then I hypothesize that female household heads would see changes in consumption through the direct effects of mobile phones and *not* through their effect on their bargaining power. Female household heads would not see changes in their relative weight Z because they occupy the primary position in the bargaining equation. It is possible for mobile phones to directly affect consumption through increasing these female heads' access to new information and resources, through enabling substitution of risky travelling, or through enabling more efficient time management and balancing of family responsibilities. Overall, this background supports my exploration of how women's mobile phone access may work out in a different household bargaining system.

This paper repeats the OLS and 2SLS instrumental variables regressions on the total consumption and consumption share variables regressed on the interaction between the dummy variable for women's mobile phone access and the dummy variable for a female household head. Since the female household head used to be one of the six instruments in the original 2SLS model, this model has only five instruments. The coefficient on the women's mobile phone access dummy alone still has the same interpretation: how total consumption or the consumption shares are different for households with at least one woman accessing a mobile phone. The coefficient on the interaction term would show the additional effect of women's mobile phone access for the 1,997 households with a female household head. Thus, the total effect of women's mobile phone access on household consumption for homes with female heads is the sum of those coefficients.

Table 16: OLS Female Head Interaction

	(1) Total Consumption	(2) Per Capita Consumption	(3) Food	(4) Education	(5) Finances	(6) Transportation	(7) Energy	(8) Household Goods	(9) Jewelry	(10) Discretionary	(11) Care Goods	(12) Intoxicants and Processed Foods
Any Woman with Mobile Access=1	-0.0102	-0.00557	-0.00216	0.00215	0.00433***	-0.00200	-0.00557***	0.000274	0.00103	-0.000245	0.00371	-0.00152*
	(-0.71)	(-0.38)	(-0.53)	(0.94)	(3.45)	(-0.92)	(-3.15)	(0.19)	(0.80)	(-0.09)	(1.14)	(-1.73)
Female Household Head=1	-0.0599**	0.0625*	0.0148*	-0.00559*	0.000946	0.00297	-0.00219	-0.00670***	-0.00125	-0.00171	0.0103	-0.0116***
	(-2.28)	(1.87)	(1.77)	(-1.82)	(0.41)	(0.98)	(-0.71)	(-3.22)	(-0.66)	(-0.48)	(1.23)	(-7.32)
Any Woman with Mobile Access=1 × Female Household Head=1	0.00854	-0.0327	-0.0140	0.00834**	-0.000134	-0.0123***	0.00916**	0.00133	0.00809**	0.000133	-0.00115	0.000611
	(0.23)	(-0.74)	(-1.38)	(2.03)	(-0.04)	(-3.00)	(2.31)	(0.45)	(2.03)	(0.02)	(-0.12)	(0.35)
Household Owns Cell Phone	0.0996***	0.0612***	-0.00936*	-0.000726	-0.00112	-0.00161	0.00678***	0.00162	0.000776	0.00506^{+}	0.00110	-0.00251*
	(5.73)	(3.44)	(-1.84)	(-0.32)	(-0.95)	(-0.87)	(2.89)	(1.05)	(0.68)	(1.52)	(0.22)	(-1.88)
Total Household Assets (Index 0-33)	0.0491***	0.0508***	-0.00849***	0.00272***	0.00160***	0.00459***	0.000253	0.00147***	0.000648***	0.000663***	-0.00259***	-0.000867***
	(29.73)	(29.03)	(-19.12)	(10.09)	(9.70)	(16.76)	(1.30)	(7.85)	(4.60)	(2.60)	(-6.41)	(-7.83)
Highest Adult Education in a Household	0.00496***	0.000179	-0.00176***	0.00144***	0.000593***	0.000520**	-0.0000176	-0.000197	-0.000152	-0.000557*	0.000604^{+}	-0.000464***
	(3.49)	(0.12)	(-4.43)	(4.85)	(4.59)	(2.35)	(-0.10)	(-1.22)	(-1.12)	(-1.92)	(1.63)	(-4.89)
Household Ratio of Adults:Children	-0.0410***	-0.111***	0.00506*	0.00611***	0.00393***	0.00219^{+}	-0.00241**	-0.00244***	-0.00344***	-0.00971***	-0.0000262	0.000746
	(-4.23)	(-10.59)	(1.84)	(4.39)	(4.27)	(1.45)	(-2.17)	(-3.06)	(-4.97)	(-4.45)	(-0.01)	(1.29)
Size of Household	0.0732***	-0.0932***	0.00467***	0.000662^{+}	-0.00197***	-0.00140***	-0.00318***	-0.000224	-0.000211	-0.000629	0.00227***	0.00000585
	(26.62)	(-28.52)	(5.90)	(1.49)	(-6.39)	(-3.17)	(-9.99)	(-0.86)	(-0.77)	(-1.17)	(3.28)	(0.05)
Log of Total HH Income	0.112***	0.104***	-0.00551***	-0.00143	0.00453***	0.00654***	-0.00388***	0.000792	0.00317***	0.00507***	-0.0100***	0.000738^{+}
	(13.52)	(12.16)	(-2.59)	(-1.08)	(6.02)	(5.99)	(-4.49)	(0.92)	(3.33)	(3.89)	(-5.46)	(1.47)
Village: Food for Work Program	0.0208	0.0155	-0.0106**	-0.00423*	0.00196	0.00393	0.00254	0.00362**	0.000737	0.00705**	-0.00555	0.000568
	(0.94)	(0.69)	(-1.98)	(-1.91)	(1.10)	(1.38)	(1.12)	(2.03)	(0.50)	(2.00)	(-1.41)	(0.50)
Village: MG-NREGA	-0.0157	-0.0293	-0.0000892	0.00734^{+}	0.00275	-0.00720	-0.0130**	0.00547**	-0.000160	0.00589	-0.00297	0.00194
	(-0.34)	(-0.57)	(-0.01)	(1.57)	(0.85)	(-1.00)	(-2.31)	(2.14)	(-0.04)	(1.11)	(-0.27)	(1.09)
Village: Other Govt Employment Guarantee	-0.0441**	-0.0437**	0.00258	-0.00135	0.00159	-0.000805	0.00169	0.00144	0.000530	0.00598*	-0.00972**	-0.00193*
	(-2.02)	(-1.98)	(0.48)	(-0.64)	(0.96)	(-0.35)	(0.59)	(0.84)	(0.31)	(1.79)	(-2.43)	(-1.83)
Village: Mahila Mandal (Women's Assoc.)	-0.00755	-0.0137	-0.00890*	0.000131	0.00120	-0.000990	0.00816***	0.000986	-0.00270**	0.000710	0.000601	0.000793
	(-0.40)	(-0.72)	(-1.83)	(0.06)	(0.87)	(-0.47)	(3.42)	(0.62)	(-2.10)	(0.25)	(0.16)	(0.78)
Electricity Access Hours/Day	-0.00584***	-0.00633***	0.000254	-0.000571***	-0.0000339	-0.000148	0.000541***	0.000126	-0.0000594	0.000494*	-0.000919***	0.000316***
	(-4.04)	(-4.32)	(0.72)	(-3.48)	(-0.30)	(-0.94)	(2.78)	(0.94)	(-0.62)	(1.83)	(-3.28)	(4.13)
Any Female Name on Household Papers	-0.0239	-0.0177	0.00439	0.0000555	0.00149	-0.00111	-0.00409*	0.00434*	0.00238	-0.00744**	-0.000969	0.000954
	(-1.30)	(-0.96)	(0.86)	(0.02)	(0.94)	(-0.47)	(-1.84)	(1.66)	(1.27)	(-2.27)	(-0.22)	(0.84)
Any Woman Works Outside Home	-0.0373**	-0.0539***	-0.00287	-0.00137	-0.00201	-0.00440**	0.00220	0.00213	-0.00427***	0.00387	0.00540	0.00132
	(-2.26)	(-3.18)	(-0.60)	(-0.56)	(-1.39)	(-2.30)	(1.18)	(1.28)	(-2.59)	(0.70)	(1.36)	(1.31)
Constant	9.063***	8.563***	0.630***	0.00547	-0.0580***	-0.0596***	0.161***	0.000803	-0.0289***	-0.00744	0.323***	0.0337***
	(93.60)	(84.52)	(24.57)	(0.39)	(-6.75)	(-4.53)	(14.84)	(0.09)	(-2.89)	(-0.54)	(14.83)	(6.38)
Observations	16447	16447	16447	16447	16447	16447	16447	16447	16447	16447	16447	16447

All regressions include region control variables.
t statistics in parentheses
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: 2SLS Female Head Interaction

	(1) Total Consumption	(2) Per Capita Consumption	(3) Food	(4) Education	(5) Finances	(6) Transportation	(7) Energy	(8) Household Goods	(9) Jewelry	(10) Discretionary	(11) Care Goods	(12) Intoxicants and Processed Foods
Any Woman with Mobile Access=1	-0.0858	0.108	0.678	-0.0934	0.00371	-0.224	-0.0168	-0.0390	-0.0887	-0.263	0.0622	-0.0200
	(-0.10)	(0.11)	(0.76)	(-0.61)	(0.04)	(-0.72)	(-0.17)	(-0.50)	(-0.74)	(-0.74)	(0.37)	(-0.38)
Female Household Head=1	1.070	2.758	5.166	-0.744	-0.0372	-1.582	-0.222	-0.359	-0.676	-1.851	0.505	-0.201
	(0.15)	(0.35)	(0.72)	(-0.60)	(-0.05)	(-0.64)	(-0.29)	(-0.59)	(-0.70)	(-0.64)	(0.38)	(-0.49)
Any Woman with Mobile Access=1 × Female Household Head=1	-3.916	-6.297	-7.558	1.304	0.377	2.466	0.0879	0.395	0.846	2.824	-0.906	0.165
	(-0.35)	(-0.50)	(-0.68)	(0.67)	(0.36)	(0.64)	(0.07)	(0.42)	(0.57)	(0.63)	(-0.42)	(0.25)
Household Owns Cell Phone	0.214	0.157	-0.144	0.0112	-0.0108	0.0391	0.0125	0.0108	0.0227	0.0571	-0.000711	0.00246
	(1.30)	(0.86)	(-0.86)	(0.39)	(-0.69)	(0.67)	(0.66)	(0.73)	(0.98)	(0.87)	(-0.02)	(0.24)
Total Household Assets (Index 0-33)	0.0500***	0.0518***	-0.00721^{+}	0.00259***	0.00169***	0.00460***	-0.0000596	0.00119***	0.000463	0.000410	-0.00267***	-0.00100***
	(12.00)	(10.16)	(-1.55)	(3.01)	(3.93)	(2.86)	(-0.16)	(3.04)	(0.74)	(0.23)	(-3.15)	(-4.27)
Highest Adult Education in a Household	0.0136	0.00816	-0.00876	0.00195	-0.000326	0.00225	0.000737	0.000575	0.00110	0.00184	0.000684	-0.0000462
	(1.38)	(0.75)	(-0.91)	(1.18)	(-0.33)	(0.68)	(0.64)	(0.66)	(0.77)	(0.51)	(0.36)	(-0.08)
Household Ratio of Adults:Children	-0.0519	-0.1000	0.0608	-0.00318	0.00541	-0.0124	-0.00559	-0.00784	-0.0119	-0.0308	0.00634	-0.000879
	(-0.67)	(-1.14)	(0.76)	(-0.23)	(0.73)	(-0.45)	(-0.66)	(-1.20)	(-1.12)	(-0.95)	(0.42)	(-0.20)
Size of Household	0.0724***	-0.0883***	0.00730	0.000586	-0.00192***	-0.00250	-0.00314***	-0.000265	-0.000544	-0.00158	0.00240**	-0.000345
	(12.87)	(-13.20)	(1.25)	(0.59)	(-3.39)	(-1.27)	(-5.30)	(-0.52)	(-0.69)	(-0.68)	(2.24)	(-1.12)
Log of Total HH Income	0.126***	0.123***	-0.00305	-0.00323	0.00283*	0.00479	-0.00288^{+}	0.00116	0.00335	0.00323	-0.00719*	0.000997
	(7.40)	(6.25)	(-0.17)	(-0.99)	(1.67)	(0.81)	(-1.51)	(0.70)	(1.38)	(0.47)	(-1.77)	(0.90)
Village: Food for Work Program	0.0132	0.0207	0.0239	-0.00839	0.00343	-0.00701	0.000212	0.000513	-0.00402	-0.00477	-0.00290	-0.000949
	(0.25)	(0.36)	(0.47)	(-0.99)	(0.69)	(-0.38)	(0.04)	(0.11)	(-0.57)	(-0.24)	(-0.30)	(-0.33)
Village: MG-NREGA	0.0654	0.0756	0.0483	-0.00342	-0.00414	-0.0232	-0.0119	0.00422	-0.00439	-0.0123	0.00510	0.00169
	(0.61)	(0.60)	(0.42)	(-0.18)	(-0.44)	(-0.60)	(-1.09)	(0.46)	(-0.29)	(-0.27)	(0.22)	(0.27)
Village: Other Govt Employment Guarantee	-0.0644**	-0.0696**	-0.00599	-0.000232	0.00267	0.00239	0.00271	0.00149	0.000599	0.00843	-0.0104^{+}	-0.00169
	(-2.12)	(-2.03)	(-0.19)	(-0.04)	(0.93)	(0.22)	(0.73)	(0.54)	(0.14)	(0.69)	(-1.59)	(-0.93)
Village: Mahila Mandal (Women's Assoc.)	0.00275	0.00133	-0.00183	-0.00153	0.0000618	-0.00401	0.00859***	0.00107	-0.00229	-0.00366	0.00254	0.00105
	(0.10)	(0.04)	(-0.06)	(-0.29)	(0.02)	(-0.40)	(2.70)	(0.38)	(-0.55)	(-0.31)	(0.42)	(0.63)
Electricity Access Hours/Day	-0.00449	-0.00559	-0.00242	-0.000254	-0.000129	0.000715	0.000531	0.000296	0.000362	0.00153	-0.00102	0.000393*
	(-1.20)	(-1.32)	(-0.60)	(-0.39)	(-0.39)	(0.51)	(1.17)	(0.91)	(0.68)	(0.95)	(-1.33)	(1.76)
Any Female Name on Household Papers	0.0114	-0.0152	-0.0951	0.00763	-0.00629	0.0254	0.00512	0.0131	0.0197	0.0250	-0.00234	0.00789
	(0.08)	(-0.10)	(-0.70)	(0.32)	(-0.46)	(0.54)	(0.35)	(1.13)	(1.04)	(0.47)	(-0.09)	(1.03)
Any Woman Works Outside Home	0.0213	-0.0143	-0.0451	-0.00113	-0.0103	0.00496	0.00940	0.00791	0.00583	0.0171	0.00483	0.00659
	(0.29)	(-0.17)	(-0.62)	(-0.09)	(-1.37)	(0.20)	(1.04)	(1.19)	(0.55)	(0.59)	(0.32)	(1.40)
Constant	8.759***	8.050***	0.313	0.0714	-0.0284	0.0502	0.158***	0.0155	0.00345	0.119	0.256**	0.0419
	(16.10)	(12.83)	(0.57)	(0.73)	(-0.54)	(0.27)	(2.78)	(0.34)	(0.05)	(0.53)	(2.29)	(1.31)
Observations	15477	15477	15477	15477	15477	15477	15477	15477	15477	15477	15477	15477
F	129.5	71.58	1.966	4.503	11.77	3.867	10.28	4.473	1.702	0.846	6.731	8.245

t statistics in parentheses
All regressions include region control variables.
$^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16 presents the results from the OLS regression on the interaction term. It suggests that relative to households headed by men but maintaining women's mobile phone access, households with a female head and women's mobile phone access have higher shares of education, energy, and jewelry spending and lower shares of transportation spending, statistically significant to the 1% or 5% level. These effects on education, transportation, and jewelry shares are supported by the household bargaining literature on consumption allocations by gender but the energy share effects warrant some further discussion. The energy share consists of spending on kerosene, telephone, cable, inter net, electricity, and household fuel. Most apparently, this could refer to an increase in telephone charges since a female household head may not restrict household members' mobile phone access as much as literature has found that male household

heads are capable of (Barboni et al. 2018, Kruse et al. 2019, Roessler et al. 2018, Rotondi et al. 2020). Of note in contrast to the original OLS results from Table 7 are the conflicting signs on the mobile phone access coefficient and the interaction term coefficient for the energy share. Reasons why these conflicting results appear can be speculated, but it seems to be as a result of the differing household bargaining dynamics at play. Further investigation of changes in energy consumption and household bargaining are warranted, such as Malhotra, Kanesathasan, and Patel 2012. Overall, the statistically significant effects on the interaction coefficient may indicate that the original OLS results of Table 7 overestimated the effect of mobile phones on consumption allocations, identifying some of the variation that should have been attributed to the greater household bargaining power maintained by a female household head. Still, no statistically significant effects were found on the interaction term for logged total consumption or logged per capita consumption, consistent with the original statistically insignificant OLS results in Table 8.

These OLS results face the same limitations as the original OLS results: there is a reasonable concern for reverse causality between some household bargaining variables and women's mobile phone access since it may also be both a factor in and result of household bargaining. This similarly motivates the 2SLS instrumental variables strategy on the interaction between the women's mobile phone access dummy and the female household head dummy. To reiterate, this specification tests how mobile phone access affects the consumption outcomes of households with and without a female household head through changing their household bargaining. There were no statistically significant effects found in any of the 2SLS regressions on women's mobile phone access, presence of a female household head, or their interaction. This test suggests that relative to male-headed households, women's mobile phone access does not have a statistically significant effect on any consumption outcome through its impact on women's bargaining power. This confirms my hypothesis that mobile phones would show only a direct impact on consumption through the three aforementioned channels. A woman has the primary bargaining power position in female-headed households, so access to a mobile phone affects *what she bargains for* due to the three channels of information access, social networking, and efficiency of activity from reduced travel. Mobile phones would not affect *how strongly she can bargain* for those preferences, as suggested by Table 17.

6 Conclusion

This paper finds evidence that households with at least one woman accessing a mobile phone allocate household consumption differently than those that do not. Women's mobile phone access seemed to decrease overall household consumption by roughly 28% between these groups of households holding many village, household, and bargaining power characteristics constant. Women's mobile phone access also seemed to increase the consumption shares of food and household finances but decrease the consumption share of transportation. Following a robustness check incorporating whether the household's bargaining model has a male or female primary actor, or rather a male or female household head, this paper finds that mobile phones do have a direct effect on consumption decisions separate from its indirect effect on women's household bargaining power. The household bargaining model and literature on mobile phone use among rural Indian women revealed a number of ways women's bargaining power could increase in the household. Of particular importance were the channels for reducing transportation costs, facilitating information gathering, and social networking with friends and family. These channels could contribute to the savings behavior potentially exhibited by the decreased total consumption, but further research on the various ways rural Indian households save is necessary to confirm this finding.

While these results expand the relevant areas of literature, more robust statistical inquiry is still needed. Precise data on women's ownership of, access to, technical capabilities regarding, and preferred usages of mobile phones could shed light on this important tool for development. Another avenue for inquiry is the use of panel data to see how new access to mobile phones for women changes consumption decisions in the short-run or long-run. As the strands of household bargaining and women's mobile phone access come together with time, perhaps creative instrumental variable strategies can uncover more about how intra-household decision making is complicated by emerging technology. The value of qualitative research in this area cannot be understated. In-depth surveys provided the personal insight that this paper's theory rests on, contributing to an understanding of *how* mobile phones change women's lives, households' status, and communities in rural India.

The savings behavior and increased nutritional concern of households with greater women's mobile access point to greater development in rural India. The economic, infrastructural, and social barriers must be engaged

with from multiple perspectives so gains like these can be realized by more families. As mobile phone access continues to grow in rural India, the successes seen in Africa with mobile money and nudging technology become more likely. Overall, mobile phones can change the horizon on many fronts: human development, poverty eradication, and gender equality.

7 References

Abor, Joshua Yindenaba, Mohammed Amidu, and Haruna Issahaku (July 3, 2018). "Mobile Telephony, Financial Inclusion and Inclusive Growth". In: *Journal of African Business* 19.3. Publisher: Routledge eprint: https://doi.org/10.1080/15228916.2017.1419332, pp.430–453. ISSN: 1522-8916. DOI: 10.1080/ 15228916.2017.1419332. URL: https://doi.org/10.1080/15228916.2017.1419332 (visited on 02/02/2022).

Aker, Jenny C. and Isaac M. Mbiti (Sept. 2010). "Mobile Phones and Economic Development in Africa". In: *Journal of Economic Perspectives* 24.3, pp. 207–232. ISSN: 0895- 3309. DOI: 10.1257/jep.24.3.207. URL: https://www.aeaweb.org/articles?id=10.1257/jep.24.3.207 (visited on 02/04/2022).

Bair, Sabrine and Ahmed Tritah (2019). "Mobile Money and Inter-Household Financial Flows: Evidence from Madagascar". In: *Revue economique ´* 70.5. Publisher: Sciences Po University Press, pp. 847–872. ISSN: 0035-2764. URL: https://www.jstor.org/stable/26762992 (visited on 02/02/2022).

Baland, Jean-Marie and Roberta Ziparo (2018). "Intra-Household Bargaining in Poor Countries". In: *Towards Gender Equity in Development*. Oxford: Oxford University Press. ISBN: 978-0-19-882959-1. DOI: 10.1093/oso/9780198829591.003.0004. URL: https://oxford.universitypressscholarship.com/10.1093/oso/9780198829591.001.0001/oso-9780198829591-chapter-4 (visited on 02/06/2022).

Barboni, Giorgia et al. (2018). *A Tough Call: Understanding Barriers to and Impacts of Women's Mobile Phone Adoption in India*. Harvard Kennedy School Evidence for Policy Design. URL: https://epod.cid.harvard.edu/sites/default/files/2018-10/A_Tough_Call.pdf (visited on 02/04/2022).

Beuermann, Diether W., Christopher McKelvey, and Renos Vakis (Nov. 1, 2012). "Mobile Phones and Economic Development in Rural Peru". In: *The Journal of Development Studies* 48.11. Publisher: Routledge eprint: https://doi.org/10.1080/00220388.2012.709615, pp. 1617–1628. ISSN: 0022-0388. DOI: 10 . 1080 / 00220388 . 2012 . 709615. URL: https://doi.org/10.1080/00220388.2012.709615 (visited on 02/03/2022).

Biswas, Raaj Kishore et al. (July 3, 2021). "Exposure of mobile phones and mass media in maternal health services use in developing nations: evidence from Urban Health Survey 2013 of Bangladesh". In: *Contemporary South Asia* 29.3. Publisher: Routledge eprint: https://doi.org/10.1080/09584935.2020.1770698, pp. 460–473. ISSN: 0958-4935. DOI: 10.1080/09584935.2020.1770698. URL: https://doi.org/10.1080/09584935.2020.1770698 (visited on 02/04/2022).

Doss, Cheryl (2005). *The Effects of Intrahousehold Property Ownership on Expenditure Patterns in Ghana*. URL: https://www.econbiz.de/Record/the-effects-of-intrahousehold-property-ownership-on-expenditure-patterns-in-ghana-doss-cheryl/10003325143 (visited on 03/22/2022).

Duflo, Esther (2003). "Grandmothers and Granddaughters: Old Age Pension and Intra-Household Allocation in South Africa". In: *World Bank Economic Review* 17. ISSN: 1.

Duflo, Esther and Christopher Udry (2004). *Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices*. Working Paper 10498. National Bureau of Economic Research. URL: https://www.nber.org/papers/w10498.

Fafchamps, Marcel, Bereket Kebede, and Agnes Quisumbing (2009). *Intrahousehold Welfare in Rural Ethiopia*. URL: https://web.stanford.edu/~fafchamp/ethwelf.pdf (visited on 03/21/2022).

Goodman, J (2005). "Linking Mobile Phone Ownership and Use to Social Capital in Rural South Africa and Tanzania". In: *Africa: The Impact of Mobile Phones*. Ed. by D Coyle. Newbury, UK: Vodafone.

Group, Telenor (2016). *Giving Women a Phone - and a Voice*. URL: http://advertisementfeature.cnn.com/2016/telenor/giving-women-a-phone/.

Hahn, Hans and Ludovic Kibora (Mar. 1, 2008). "The Domestication of the Mobile Phone: Oral Society and New ICT in Burkina Faso". In: *The Journal of Modern African Studies* 46, pp. 87–109. DOI: 10.1017/S0022278X07003084.

Hyde, Elizabeth, Margaret E Greene, and Gary L Darmstadt (2020). "Time poverty: Obstacle to women's human rights, health and sustainable development". In: *Journal of Global Health* 10.2, p. 020313. ISSN: 2047-2978. DOI: 10.7189/jogh.10.020313. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7688061/.

Jack, William and Tavneet Suri (Jan. 2014). "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution". In: *American Economic Review* 104.1, pp. 183–223. ISSN: 0002-8282. DOI: 10.1257/aer.104.1.183. URL: https://www.aeaweb.org/articles?id=10.1257/aer.104.1.183 (visited on 02/04/2022).

Kruse, Clemens et al. (Oct. 9, 2019). "Barriers to the Use of Mobile Health in Improving Health Outcomes in Developing Countries: Systematic Review". In: *Journal of Medical Internet Research* 21.10. Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada, e13263. DOI: 10.2196/13263. URL: https://www.jmir.org/2019/10/e13263 (visited on 02/04/2022).

Labonne, Julien and Robert Chase (2009). *The Power of Information: The Impact of Mobile Phones on Farmers' Welfare in the Philippines*. Policy Research Working Paper 4996. URL: https://documents1.worldbank.org/curated/en/132511468297548935/pdf/WPS4996.pdf.

Malhotra, Anju, Anjala Kanesathasan, and Payal Patel (2012). *Connectivity: How Mobile Phones, Computers, and the Internet Can Catalyze Women's Entrepreneurship, India: A Case Study*. International Center for Research on Women Report. URL: https://www.icrw.org/wp-content/uploads/2016/10/Connectivity-how-mobile-phones-computers-and-the-internet-can-catalyze-womens-entrepreneurship.pdf.

Mehrotra, Nilika (2004). "Gold and Gender in India : Some Observations from South Orissa". In: *Indian Anthropologist* 34.1. Publisher: Indian Anthropological Association, pp. 27–39. ISSN: 0970-0927. URL: https ://www.jstor.org/stable/41919946 (visited on 03/20/2022).

Mehta, Balwant Singh (2013). *Capabilities, Costs, Networks and Innovations: Impact of Mobile Phones in Rural India*. Institute for Human Development Working Paper 29. URL: https://assets.publishing.service.gov.uk/media/57a08a2740f0b652dd0005ba/ctg-wp-2013-29.pdf.

Mehta, Balwant Singh and Nidhi Mehta (2014). "ICT and Socio-Economic Empowerment of Rural Women: Case of Mobile Phone in India". In: *Knowledge Horizons - Economics* 6, pp. 103–112. ISSN: 6. URL: http://www.orizonturi.ucdc.ro/arhiva/2014_khe_6_pdf4/balwant.pdf.

Mohan, Diwakar et al. (July 20, 2020). "Does having a mobile phone matter? Linking phone access among women to health in India: An exploratory analysis of the National Family Health Survey". In: *PLOS ONE* 15.7. Publisher: Public Library of Science, e0236078. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0236078. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0236078 (visited on 01/22/2022).

Mohapatra, Sandeep and Leo Simon (2017). "Intra-Household Bargaining over House-hold Technology Adoption". In: *Review of Economics of the Household* 15.4, pp. 1263–1290. URL: https://www.proquest.com/docview/1957691478?pq- origsite=gscholar&fromopenview=true.

Moore, Charity Troyer and Rashi Sabherwal (2017). Can India Answer the Call? Adress-ing the Gender Gap in Mobile Phone Access. The Indian Express. URL: https://indianexpress.com/article/technology/opinion-technology/can-india-answer-the-call-addressing-the-gender-gap-in-mobile-phone-access-4995705/.

Munyegera, Ggombe Kasim and Tomoya Matsumoto (Mar. 1, 2016). "Mobile Money, Remittances, and Household Welfare: Panel Evidence from Rural Uganda". In: World De-velopment 79, pp. 127–137. ISSN: 0305-750X. DOI: 10.1016/j.worlddev.2015.11.006. URL: https://www.sciencedirect.com/science/article/pii/S0305750X15002880 (visited on 02/04/2022).

Muto, Megumi and Takashi Yamano (2009). "The Impact of Mobile Phone Coverage Expansion on Market Participation: Panel Data Evidence from Uganda". In: *World Development* 37.12. Publisher: Elsevier, pp. 1887–1896. ISSN: 0305-750X. URL: https://econpapers.repec.org/article/eeewdevel/v_3a37_3ay_3a2009_3ai_3a12_3ap_3a1887-1896.htm (visited on 02/04/2022).

Overå, Ragnhild (July 1, 2006). "Networks, distance, and trust: Telecommunications Development and changing trading practices in Ghana". In: *World Development* 34.7, pp. 1301–1315. ISSN: 0305-750X. DOI: 10.1016/j.worlddev.2005.11.015. URL:https://www.sciencedirect.com/science/article/pii/S0305750X06000660 (visited on 02/04/2022).

Pajaron, Marjorie (2011). *The Impact of Gender on the Intrahousehold Allocations of Remittances of Filipino Migrant Workers*. URL: https://epc2012.princeton.edu/papers/120980 (visited on 03/21/2022).

Pande, Rohini and Simone Schaner (2017). *The Mobile Phone Gender Gap: Why Does It Matter and What Can We Do?* The Indian Express. URL: https://indianexpress.com/article/technology/the- mobile- phone-gender- gap- why- does- it-matter-and-what-can-we-do/.

31. Riley, Emma (Nov. 1, 2018). "Mobile money and risk sharing against village shocks". In: *Journal of Development Economics* 135, pp. 43–58. ISSN: 0304-3878. DOI: 10.1016/j.jdeveco.2018.06.015. URL: https://www.sciencedirect.com/science/article/pii/S0304387818304413 (visited on 02/04/2022).

Roessler, Philip et al. (2018). "Mobile-Phone Ownership Increases Poor Women's House-hold Consumption: A Field Experiment in Tanzania". In: Evidence in Governance and Politics Meeting. Nairobi, Kenya. URL: https ://www.icrw.org/wp-content/uploads/2019/02/Nielson-Dan_Paper.pdf (visited on 02/02/2022).

33. Rotondi, Valentina et al. (June 16, 2020). "Leveraging mobile phones to attain sustainable development". In: *Proceedings of the National Academy of Sciences* 117.24. Publisher: National Academy of Sciences Section: Social Sciences, pp. 13413–13420. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1909326117. URL: https://www.pnas.org/content/117/24/13413 (visited on 01/22/2022).

Scott, Kerry et al. (Sept. 1, 2021). "Freedom within a cage: how patriarchal gender norms limit women's use of mobile phones in rural central India". In: *BMJ Global Health* 6 (Suppl 5). Publisher: BMJ Specialist Journals Section: Original research, e005596. ISSN: 2059-7908. DOI: 10.1136/bmjgh-2021-005596. URL: https://gh.bmj.com/content/6/Suppl_5/e005596 (visited on 02/04/2022).

Sekabira, Haruna and Matin Qaim (Dec. 1, 2017). “Can mobile phones improve gender equality and nutrition? Panel data evidence from farm households in Uganda”. In: *Food Policy* 73, pp. 95–103. ISSN: 0306-9192. DOI: 10.1016/j.foodpol.2017.10.004.URL: https ://www.sciencedirect.com/science/article/pii/S0306919217303093 (visited on 02/06/2022).

Sonne, Lina (2020). “What Do We Know About Women’s Mobile Phone Access & Use? A Review of Evidence”. Working Paper No. WP-2020-03. Dvara Research Working Paper Series. URL: https://www.dvara.com/research/wp-content/uploads/2020/06/What-Do-We-Know-About-Womens-Mobile-Phone-Access-Use-A-review-of-evidence.pdf.

Souter, David et al. (2005). *The Economic Impact of Telecommunications on Rural Livelihoods and Poverty Reduction: A Study of Rural Communities in India (Gujarat), Mozambique and Tanzania*. DFID KaR Project 8347. (Commonwealth Telecommunications Organisation for UK Department for International Development. URL: https://assets.publishing.service.gov.uk/media/57a08c9f40f0b652dd001446/3943_R8347-Econ-Impact-TeleCom-Rural-Livelihoods.pdf (visited on 02/03/2022).

Stark, Laura (Apr. 2, 2020). “Women, Empowerment, and Mobile Phones in the Developing World”. In: *The Oxford Handbook of Mobile Communication and Society*. ISBN: 978-0-19-086438-5. DOI: 10.1093/oxfordhb/9780190864385.013.35. URL: https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780190864385.001.0001/oxfordhb-9780190864385-e-35 (visited on 02/04/2022).

Suri, Tavneet and William Jack (Dec. 9, 2016). “The long-run poverty and gender impacts of mobile money”. In: *Science*. Publisher: American Association for the Advancement of Science. DOI: 10.1126/science.aah5309. URL: https://www.science.org/doi/abs/10.1126/science.aah5309 (visited on 02/03/2022).

Tankari, Mahamadou Roufahi (June 2018). “Mobile Phone and Households’ Poverty: Evidence from Niger”. In: *Journal of Economic Development* 43.2. Num Pages: 67-84 Place: Seoul, South Korea Publisher: The Economic Research Institute, Chung-Ang University, pp. 67–84. ISSN: 0254-8372. URL: https://www.proquest.com/docview/2131783890/abstract/DAD8D47F707C4D70PQ/1 (visited on 02/02/2022).

41. Wang, Shing-Yi (Mar. 1, 2014). “Property rights and intra-household bargaining”. In: *Journal of Development Economics* 107, pp. 192–201. ISSN: 0304-3878. DOI: 10.1016/j.jdeveco.2013.12.003. URL: https://www.sciencedirect.com/science/article/pii/S0304387813001752 (visited on 03/22/2022).

Does Policing Work? Evaluating the Impact of Police Budgets on Crime in US Urban Cities

Abstract

Increasing police funding as a method to reduce crime has been a salient political issue for decades that has recently become increasingly contentious with the Black Lives Matter protests of summer 2020. This report uses statistical evidence to evaluate the relationship between crime rates and a recently published police spending report on major urban cities in the United States. Prior literature has yielded mixed results and displayed that police spending may be more reactive to crime than preventative. We employed a variety of different empirical frameworks, which yielded results displaying few significant relationships between our variables. Population was the only variable consistently positive and significant with crime, while police spending and other factors were rendered more ambiguous. Overall, our report, as does much of the literature, demonstrates that our data did not support the idea that increased police spending decreased crime.

Content

1 Introduction

The recent killings of George Floyd, Brianna Taylor, Ahmaud Arbury, and other African American citizens set the country ablaze during the summer of 2020. In response, mobilization of the Black Lives Matter protests and other grassroots organizations brought policing to the forefront of political discourse. Slogans of "defund the police" and "abolish the police" were chanted as protests erupted across the country. Proponents of racial justice emphasized law enforcement as a form of systemic racism that evolved from slave catchers and sheriffs into racially targeted policies such as "stop and frisk" and the "war on drugs" (Zerkel, 2020). These proponents argue this legacy has contributed to the problem of African-Americans being incarcerated at over five times the rate of white people (Zerkel, 2020). In addition to the history of policing, activists asserted that defunding the police would allow local governments to divert more money toward social services and away from the militarization of policing. This includes funding for town centers, mental health institutions, educational facilities, and other social programs that would be more effective in strengthening local communities. On the other side of the spectrum, many suggest that defunding the police is not an effective long-term solution. Many surveys demonstrate that residents of low-income, high-crime communities support more police intervention to serve as protection against internal violence in their neighborhoods. Data supporting these concerns indicate a positive correlation between increased police presence and reduced robbery and other certain types of crime in urban communities like Oakland, California (Lofstrom and Martin, 2022). Furthermore, select cities with reduced police budgets following 2020 have faced unsatisfactory outcomes for various reasons (Goodman, 2021).

In this report, we will utilize empirical and quantifiable data to evaluate this political issue, examining the links between police funding and local crime. We will use data compiled by *Vera*, for information on the "percent of city funds spent on policing" and the "city [dollars per resident for police] for the 2020 fiscal year of October 1, 2019, to September 30" (Vera, 2020)[1].These "per capita" metrics will allow us to examine a specific number of major urban cities in the country, without having to weigh the overall policing budget of one city over another. This study will be the first to use this new data compiled by Vera, to analyze the relationship between modern police spending and crime in various metropolitan areas. For the crime statistics, we will use data

1 This study included 72 cities, 70 of which we found corresponding crime data to include in our analysis.

from the 2019 FBI calendar year for each of these corresponding cities[2] . The data encompasses cities with a cumulative population of over 64 million. We will evaluate both the overall crime rights as well as several specific categories including murder, robbery, assault, and burglary (FBI, 2020). At the time we started writing this report, the 2020 FBI data had not been published, however, there are still 3 overlapping months of data and it is unlikely that police budgets in these cities shifted dramatically from 2019 to 2020. we started writing this report, the 2020 FBI data had not been published, however, there are still 3 overlapping months of data and it is unlikely that police budgets in these cities shifted dramatically from 2019 to 2020.

The goal of this report will be to analyze if there is a relationship between police spending and crime rates, based on empirical evidence. While these results have significant political implications, crime and policing are not expected to be linear. As displayed by prior research as well as the existence of crime in contemporary society, policing is not a solution to eradicate crime, but rather a tool to mitigate its occurrences and repercussions. Additionally, for crime to be reported within crime statistics requires that certain illegal activity be observed and reported. The existence of unsolicited and unknown illegal activity brings into question the reliability of crime statistics as a proper measure of public safety. Additional considerations outside the framework of this analysis could be to evaluate the behavioral effects of increased policing, focus on the correlation between spending on social programs and policing, or use a time-series analysis to evaluate change over the years.

2 Literature Review

Research on a correlation between police and crime has transpired for decades. The first prominent publication on the issue was titled "Crime and Punishment: An Economic Approach" and was published by University of Chicago professor Gary Becker, who was working for the National Bureau of Economic Research at the time (1968). This paper was the first academic attempt at analyzing the deterrence effect, which is the rationale that police can effectively deter crime from taking place. Becker emphasized that the goal of his study was normative questions on what type of punishments should be utilized. However, Becker directs his focus on creating models that mathematically optimize the punishment

2 The vast majority of our cities (66/70) utilized 2019 FBI crime data. However, data for Chicago, Newark, New York, and Philadelphia was not published by the FBI, so we chose to find alternative sources (citations in bibliography).

that should be utilized. Becker, himself, stated that the paper was heavily based on theoretical models, as the "essay concentrates almost entirely on determining optimal policies to combat illegal behavior and pays little attention to the actual policies" (1968, p. 44-45). Even so, many future studies reference this initial effort and method of applying economics to law enforcement.

A subsequent study by Thomas Pouge, published in 1975, found that differences in the crime rates of cities they investigated could not be based on differences in resources devoted to law enforcement (Urban Institute, 2022). Additionally, Pougue had conflicting findings on demographics; the evidence supported a trend between crime rates and demographics through cross-sectional analysis but did not find any correlation by anlyzing a shift in demographics through time-series data between 1962 and 1967. However, Pogue concluded that there was a correlation between the nonwhite population and the crime rates, which was not explained by poverty rates. This was a controversial statement, that paved the way for further research on the issue. This study was conducted very differently than our own, but it is important to compare how current research in the area has evolved from the first scholars.

The majority of studies conducted during the 1980s differed from the results of these early scholars and did not find a correlation between these two variables. Samuel Cameron published a study in which he reviewed the theoretical frameworks and empirical data on how punishment served as a deterrent to criminal activity (1988). Although Cameron was not focused on police budgets specifically, this work was referenced by dozens of others in the field. He found that there was no correlation between punishment and crime, and interpreted this to mean that there was no correlation between the number of police to enforce these punishments and crime rates as well. A similar study by Brier and Findberg supported Cameron's analysis and found flaws with the model developed by Becker (1980).

Later on, Steven Levitt conducted a meta-analysis on how nationwide crime rates fell in the 1990s by unprecedented amounts. Levitt analyzed six factors that did not explain the decline including the strength of the economy, changing demographics, better policing strategies, gun control laws, laws allowing the carrying of concealed weapons, and increased use of capital punishment. Levitt described how studies before the 1990s tended to result in an insignificant correlation because they did not account for the endogeneity problem. That is, the increasing number

of police was a response to an increase in crime. He concluded that "the increase in police between 1991 and 2001 can account for a crime reduction of 5-6 percent across the board" (p. 418). One more step that Levitt took was running a simple cost-benefit analysis, in which he concluded that the increase in police budgets would cost around $8.4 billion, but reducing crime by 5-6% would save approximately $20-25 billion. Therefore, the cost of preventing crime was only one-third the amount of money that would be saved from the cost of crimes from occurring. In addition to the increasing the size of the police force, Levitt also found that the rising prison population, the receding crack epidemic, and the legalization of abortion also played a role in decreasing crime rates during the 1990s.

Levitt also published a paper hypothesizing that electoral cycles influence how many police are hired each year (1997). In analyzing this hypothesis, he concluded that "Increases in the size of police forces are shown to be disproportionately concentrated in mayoral and gubernatorial election years" (p. 270). However, the results also yielded interesting results about specific crime statistics. Levitt concluded that the size of police forces substantially reduced violent crime (murder, assault, etc.) but had little impact on property crime (burglary, larceny-theft, etc.). This study is very significant because the data demonstrates that police spending is a political issue that is at the front of many incumbents' thoughts. Even though politicians may not have access to and knowledge of very much data, they tend to make calculated decisions on spending based on what they believe the impact will be. In addition, our analysis is very similar to Levitt's method of using data from different cities to analyze the relationship between the police and rates of different crimes.

A critique of Levitt's work was published by Justin McCray, who describes an error in Levitt's computer computations (2002). McCray contends that this impacted Levitt's assessment of violent crime because even though there was still a negative correlation between violent crime and police spending, the significance level dramatically decreased from 0.06, which is considered significant to weaker levels of 0.61. More recently, Ming-Jen Lin used aggregate state data from 1980 to 2000 to evaluate the effect of increasing the number of police on crime (2009). Lin discovered a negative relationship between both property and violent crime by using a more complex two least-squares regression analysis; however, only property crimes were statistically significant, in contrast to Levitt's models. The uncertainties demonstrated in the previous studies affirm the importance of evaluating different types of crime statistics amongst different populations.

In more modern work, evidence has become more nuanced about what specific tasks that police are effective at. In regards to evaluating different types of crime, studies reveal that police effectiveness in their ability to prevent crime depends on the type of crime committed. For example, Dan Burges asserts that theft prevention relies almost entirely on the owner of the property because of police forces' inability to patrol all geographic areas simultaneously (2013). The role of police in this scenario is more reactive than preventative and is not likely to change with increased police funding. By considering this information, we decided that our experiment will segment crime by certain types in order to further explore the effectiveness of policing in reducing crime.

One prior study used quantitative and qualitative analysis to review the impacts of defunding the police, very similar to the approach and method we are utilizing. In this study, Stephen Rushin and Roger Michalski evaluate the police expenditures per capita in every locality across the United States (2020). They use data on underfunded policing areas and specific case studies such as Cleveland, Ohio, and Oakland, California to conclude that a lack of funding correlates with increased officer misconduct and violence. Rushin and Michalski suggest that this may be due to a lack of training, resources, and ability to oversee the behavior of officers. They make a comparison between policing and schools, suggesting that policing services should be seen as a similar public good in order to equalize funding across departments. These would be the first steps in initiating an institutional transformation to reform the police rather than defund them.

Building on years of historical scholarship, The Office of Community Oriented Policing Services (COPS), under the Department of Justice, researched the relationship between economic conditions, policing, and crime trends in a report (2012). This resource not only considers the police's impact on crime but also economic conditions and other social factors. In analyzing the Great Recession's impact on crime, COPS emphasizes a key component relative to this research: "identifying economic variables that affect crime is difficult because although there may be a correlation between the two," that does not imply direct causation (Scheider et al., 2012, p. 3). This acknowledgment highlights the complexity of policing crime as crime trends and economic conditions are contingent on local conditions (Scheider et al., 2012). Referencing scholars like Levitt, the report corroborates that the size of the police force

matters in the ability to control crime to some extent. In addition to these findings, this report indicates that most research has focused on the effects of generic changes in the size of police forces and thus does not account for how changes in "personnel, equipment, analytical resources, organization support for problem-solving," and other factors affect local crime (Scheider et al., 2012, p. 11). To address this limitation, this report will focus on changes to police budgets which encompasses changes in personnel, equipment, and technology rather than strictly focusing on the total number of officers.

On the topic of increasing budgets for the growth of the police, a study by Jonathan Mummolo of the Proceedings of the National Academy of Sciences of the United States of America found no firm evidence that SWAT teams for crime deterrence lower an agency's violent crime rate or the rates at which officers are killed or assaulted (2018). In addition, using survey data, he determined that this militarization of the police leads citizens to react negatively to their presence. This in turn results in an increased unwillingness to fund police agencies (Mummolo, 2018). The nuanced relationship between police funding, crime, and the public perception of law enforcement again highlights that there is no simple answer to increasing public safety, as it is a political issue by nature. However, by analyzing the links between police spending among other factors across a specific locality, this paper explores the ramifications of adjusting police budgets.

This relationship emphasizes the complexity of crime reduction as i creased policing is not a feasible solution to eradicate all crime. In addition, a 2000 study conducted in The Review of Economics and Statistics corroborates other reports that income inequality has little effect on property crime, but strong and robust impacts on violent crime (Kelly, 2000). This relationship was found even when controlling for the effects of poverty, race, and family consumption. This study focused specifically on the United States and has contemporary relevance as the issue of income inequality continues to persist. Again, relevant crime studies reveal the difficulty of minimizing crime by highlighting how structural changes to income distribution and education may have major implications on crime regardless of police presence.

The inability of previous research to reach a consensus on the relationship between increased policing and crime rates highlights the complexity of policing and increasing public safety. As evident by the research, crime does not follow a rational or simple path that can be predicatable

and preventative (Scheider et al., 2012). Given this reality, this paper seeks not the solution to effectively manage crime but rather explores how different municipalities' approaches to policing may affect their local crime conditions. Because local crime conditions are dependent on geography, weather, structural environments, access to education, and affluence, among many factors listed above, we will analyze additional variables such as population, income statistics, and education levels to paint the full picture of crime rates within a given city. These variables will also provide comparable regions and cities for our analysis.
Additionally, our research will focus on large metropolitan cities to mi igate the fact that crime conditions vary in different environments as demonstrated by the research presented.

3 Data Description and Methodology

For our original model, we used multiple dependent and explanatory variables to analyze the effects of crime in 70 different cities across the United States, primarily consisting of large, urban centers. Nine of these variables are crime statistics and seven were independent variables or potential factors that may impact the crime rates. The crime statistics were collected from 2019 FBI data. These statistics include violent crime, which was a sum of the subcategories murder, rape, robbery, and aggravated assault, and property crime, which was a sum of subcategories of burglary, larceny-theft, and motor vehicle theft across different cities. To account for population changes across the different cities we kept the population as a regressor to consider the effect that the size of a city may have on crime. In addition to population, we included several other variables that are utilized to examine crime and often appear in prior literature. These include variables relating to income, such as the median household income and the poverty rate, as well as education, such as the high school graduation rate, and the percentage of the population over 25 with a bachelor's degree or higher for each respective city. As studied, education is predicted to have an inverse relationship with crime (Loc ner and Moretti, 2014). Furthermore, income inequality has also been found to have effects on crime rates across municipalities (Kelly, 2000). However, the two primary regressors that we are focusing on for this paper are the amount of police funding per city resident and the percent of city funds spent on police. As mentioned previously, crime is a complex and multifaceted issue with many contributing factors and it is important to acknowledge there are factors beyond these such as specific police training methods and operations as well as social and cultural factors, which

undoubtedly impact our model and cannot all feasibly be accounted for.

Table 1 below provides the summary statistics of our data source. As evident by the table, on average the cities in the study spent about $404 per resident on policing in the 2020 fiscal year. This accounted on average for 28.7% of all city spending. With highs and lows of $840 and $184 per resident, police funding varies greatly across locales. Regarding crime statistics, on average cities experienced about 70 incidents of violent crime per 10,000 residents alongside about 337 incidents of property crime per 10,000 residents in the year. Again, the crime statistics' standard devi tion and minimum and maximum values point to differing conditions amongst cities. The table also displays some of the additional factors we will analyze such as the population, percentage of the population in poverty, high school diploma and bachelor's degree attainment rates, and median household income. With an average population for cities of 819,000, this study primarily focuses on the effects of police spending on large metropolitan cities in the United States.

Table 1: Summary Statistic[3]

VARIABLES	(1) N	(2) Mean	(3) Std. Dev	(4) Min	(5) Max
Violent Crime	68	70.71	42.54	2.780	196.5
Murder	69	1.179	1.256	0.0300	6.460
Rape	69	6.433	3.452	0.360	19.62
Robbery	70	18.40	13.18	1	81.31
Aggravated Assault	70	52.13	58.74	1.390	470.4
Property Crime	69	336.8	138.6	30.79	662.6
Burglary	70	57.20	52.91	4.040	431.1
Larceny Theft	70	235.9	101.7	22.69	469.6
Motor Vehicle Theft	68	50.35	36.71	4.040	274.5
City Funds Per Resident	70	404.1	149.1	184	840
% of City Funds Spent on Police	70	0.287	0.115	0.0500	0.640
Population	70	818,662	1.124e+06	42,958	8.337e+06
Median Income	70	61,497	15,904	32,498	119,136
Poverty Rate	69	0.163	0.0529	0.0700	0.332
% of High School Graduates	70	0.879	0.0496	0.760	0.960
% Bachelor's Degree Holders	70	0.374	0.105	0.160	0.650

3 One important note is that Table 1 is referring to crime statistics in terms of population per 10K to more easily compare and conceptualize. For our regression analysis, we used aggregate crime rates as a dependent variables, with population as an independent variable to account for the effects of the size of the city.

In addition, Figures 1 and 2 graphically display the correlation between violent crime an police spending and property crime and police spending. Despite the rationale that increased police spending would indicate increased police presence and would be correlated with lower crime rates as demonstrated in select studies, the correlations reveal the opposite (Lin, 2009). In fact, police spending in the sample was observed positively correlated with violent and property crime except in the case of property crime and police spending per resident. While this simple correlation does not imply causality between the two variables, it may point to an important factor that will affect the modeling. For one, areas with more crime tend to hire more police (Lin, 2009). Additionally, areas with larger populations may present more opportunities for crime to occur despite spending a proportional amount on police budgets. Therefore, we can not conclude that increasing police budgets is ineffective in reducing crime from this simple causality, but rather, it may elicit a factor that will affect how scholars have modeled the phenomena to mitigate this bidirectional relationship (Lin, 2009).

Figure 1: Correlations between Violent Crime and Police Spending

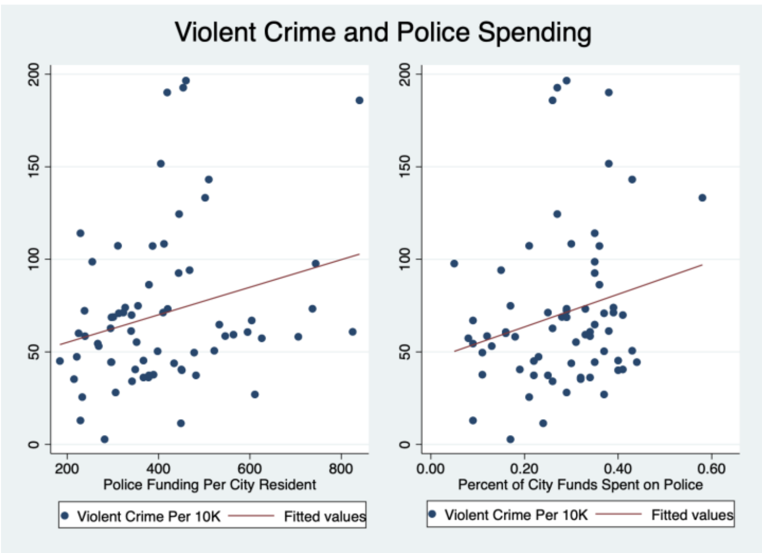


Figure 2: Correlations between Property Crime and Police Spending

To analyze the impact that police spending has on crime we first utilized an ordinary least squares (OLS) regression. This approach is backed by rigorous econometric methods demonstrating that OLS estimators are "best linear unbiased estimators" (BLUE). We ran separate regressions for violent crime, property crime, and each sub-category of crime within these categories. Based on our study of the literature, we first looked for a linear relationship between police spending and crime rates while controlling for the relevant factors discussed above. The equation below is the baseline model where we started our analysis before moving into more complex models used by previous researchers to uncover this relationship. The use of more complex models reflects both attempts to emulate relevant models studied in the field and also our response to potential hazards displayed in the baseline model.

Equation 1: OLS Model Generic Equation Formula

$$
\begin{aligned}
&\text{"}\mathit{Crime\ Statistic}\text{"} = \beta_1 + \beta_2 \text{City Police Funds Per Resident}_i + \\
&\beta_3 \text{\% City Funds Spent on } \mathit{Police}_i + \beta_4 \text{Population}_i + \beta_5 \text{Median Income}_i + \\
&\beta_6 \mathit{Poverty\ Rate}_i + \beta_7 \text{\% of High School Graduates}_i + \beta_8 \text{\% Bachelor's Degree Holders}_i + \varepsilon_i
\end{aligned}
$$

However, upon examination, the baseline model to test the linear relationship between police spending and crime is ineffective in analyzing this relationship. The Gauss-Markov Assumptions are key to OLS estition because they demonstrated that OLS estimators are unbiased and the most efficient estimators (Hallin, 2014). However, our estimators violated one of the Gauss-Markov assumptions, which was the principle of homoskedasticity (constant variance of the error term). For OLS estimators to be BLUE, they must have a constant variance of the error term, otherwise, while the coefficient estimates remain unbiased, the estimated standard errors are adversely affected which means that statistical inference can no longer be relied upon (Introductory Economics). As a result of heteroskedasticity, this model is inefficient and its inference does not provide insight into the relationship we seek to explain.

To test for heteroskedasticity, we utilized both Breush-Pagan and White Tests. For each dependent variable, based on their respective Breush-Pagan and White Test statistics, we rejected the null hypothesis that there was no heteroskedasticity for almost all of the models. The exception is the aggravated assault model, but we rejected the majority of dependent variables as detailed in appendix A[4]. In addition, for all of our models, except for the aggravated assault results, we rejected the hypothesis that results, we rejected the hypothesis that there were no omitted variables in the regressions by using a Ramsey RESET Test. These RESET tests indicated that there was at least one non-linear term in the data, which alongside models utilized in the literature motivated us to consider transforming our models.

Based on the principles of statistical inference, we decided that the generic linear OLS model we utilized is insufficient to correctly model our data. A significant portion of papers that we analyzed in the prior literature section had utilized a logarithmic model to compute elasticities between crime rates and police spending. In one article, the justification used for a logarithmic model was that using logarithmic crime rates can mitigate the problem of unreported or unrecorded crime because any measurement error is likely proportional to the true value (Ehrlic, 1996; Lin, 2009). Furthermore, previous academic literature has examined elasticities between crime and police spending. By taking the log of the dependent variable and also of terms that are not expressed in percentage terms, "the estimated coefficients are immediately interpretable as elasticities" (Kelly, 2000). Finally, not only is logging the dependent variable as well as key components consistent with the literature but also can reduce the issue of heteroskedasticity and omitted non-linear terms found previously in the models which rendered their statistical inference

4 Specifically, we only failed to reject aggravated assault in both tests, rape under the White test, and property crime and larceny under the Breush-Pagan test.

unreliable.

Equation 2: Semi-Log Linear Model

$$log(\text{"}Crime\ Statistic\text{"}) = \beta_1 + \beta_2 log(\text{City Police Funds Per Resident})_i + \beta_3\%\ \text{City Funds Spent on}\ Police_i + \beta_4 log(Population)_i + \beta_5 log(Median\ Income)_i + \beta_6 Poverty\ Rate_i + \beta_7\%\ of\ High\ School\ Graduates_i + \beta_8\%\ Bachelor's\ Degree\ Holders_i + \varepsilon_i$$

When comparing the Akaike and Bayesian Information Criteria between the generic OLS model we used and the semi-log model, we saw a dramatic decline in the number, ranging from 700-1500 in the original OLS model to roughly 100-200 in the logged model[5]. This measurement confirmed our intuition that a logarithmic model would better represent the relationship based on the nature of the transformation, backing up previous literature For this model, we also tested the goodness of fit between logging the variables "police funds per city resident," "median income," and "population" or simply logging the dependent variable crime statistic. We found that by logging all of the explanatory variables not expressed in percentage terms we achieved the best goodness of fit which was consistent with the literature and also provided us with elasticities between the explanatory variables and the dependent variables.

Variables expressed in terms of percentages such as the percentage of city funds spent on policing were not logged as they will already give an ela ticity when the dependent variable was logged.

In addition to the goodness of fit, the log transformation of the model had effects on the variance of the error term. Utilizing Breush-Pagan and White Tests, we found that these models fit the Gauss-Markov Assumption of homoskedasticity except in the case of the models with murder, rape, robbery, and motor vehicle theft as depicted in appendix B[6]. To account for the existence of heteroskedasticity in the murder, rape, robbery, and motor-vehicle theft regressions we ran these models using White (Robust) Standard Errors to obtain unbiased standard errors of the coefficient terms ("Robust Standard Errors"). While these standard errors may be larger than traditional standard errors, they will provide estimates that will allow us to have more reliable statistical inferences for those models. Also, when running RESET tests for all of these new

5 The only value outside this range was 97.3, which was the AIC for Larceny.

6 We only rejected murder and robbery for both tests, but only rape and motor vehicle theft were rejected using the BP test. However, we decided to include robust standard errors for all of these statistics due to how close these p-values were to the alpha levels in the White test.

logarithmic models, we failed to reject the hypothesis that there was an omitted variable for all of the crime statistics except motor vehicle theft. Clearly, this transformation positively affected a potential omitted variable bias and signifies an improvement in the modeling.

While the semi-log model is consistent with select previous studies, increased the goodness of fit, and also reduced heteroskedasticity, it still violates a key principle of the Gauss-Markov Assumptions which is that the regressors must be exogenous with the error term. In a standard OLS model, the regression "assumes that errors in the dependent variable are uncorrelated with the independent variable" ("Two-Stage Least-Squares Regression," 2022). Given the nature of crime and crime prevention, these variables may be endogenous in some circumstances. As discussed, changes to police spending may come as a direct response to increasing crime, and crime rates may be affected by changes to police spending (Lin, 2009). This simultaneity may be better modeled with a two-stage least squares regression which uses instrumental variables exogenous to the dependent variable to compensate for the fact that the standard OLS model may not be optimal to provide these estimates ("Two-Stage Least-Squares Regression," 2022). Therefore, for our final model, we used a two-stage least squares regression model to further explore this potentially bidirectional relationship. This model has been used consistently in this area of research by scholars to compensate for this relationship between crime and policing (Cornwell and Trumbell 1994; Kelly 2000; Lin, 2009). Lin found that using a 2SLS model as compared to OLS resulted in far more negative and statistically significant estimates concerning policing and crime (2009).

Previous literature utilized numerous instrumental variables to obtain 2SLS estimators. For instrumental variables to be valid, they must be exogenous to the independent variable (crime) and should be correlated with the endogenous variable (police spending) (Baskel, 2008). Cornwell and Trumbell argued that per capita tax revenue was a valid instrument because counties with residents who have a greater preference for law enforcement will express those preferences by voting for higher taxes (1994). Therefore, those counties will have larger police spending and presence "for reasons not directly related to the crime rate" (Cornwell and Trumbell, 1994). Furthermore, Lin argues that one year lagged state sales tax rates are the best instrument because "state revenues generated by state sales tax rates can be channeled by state transfers to local governments" hence increasing police spending and presence while being exogenous to crime rates (2009). Both studies provide statistical proof of

the validity of these instruments, therefore we employ both state sales tax rate and per capita revenues as instruments to counter our two endogenous variables, "police funding per city resident" and "percent of city funds spent on police."

Equation 3: 2SLS Equation (Second Stage Regression)[7]

$$\begin{aligned}&log(\text{"}Crime\ Statistic\text{"}) = \beta_1 + \beta_2 log(\text{City Police }\widehat{\text{Funds}}\text{ Per Resident})_i + \\ &\beta_3\%\ of\ City\ \widehat{Funds}\text{ Spent on }Police_i + \beta_4 log(Population)_i + \beta_5 log(Median\ Income)_i + \\ &\beta_6 Poverty\ Rate_i + \beta_7\%\ of\ High\ School\ Graduates_i + \\ &\beta_8\%\ Bachelor's\ Degree\ Holders_i + \varepsilon_i\end{aligned}$$

4 Results

Tables 2 and 3 display the results of the regression using the generic linear OLS equation including the coefficients and their standard errors. Coefficients with a single asterisk are significant with an alpha level of 0.1, those with two asterisks are significant with an alpha level of 0.05, and those with three asterisks are significant with an alpha level of 0.01. As expected, a positive population coefficient was statistically significant across all types of crime demonstrating the intuitive fact that cities with larger populations experience more crime. Yet, given the heteroskedasticity of the error terms and the results of the RESET test, the inference is not reliable. Furthermore, the goodness of fit indicators reveals that these results are likely not representative of the true relationships between crime and the explanatory variables.

Table 2: Violent Crime Generic OLS Equation Results

VARIABLES	(1) Violent Crime	(2) Murder	(3) Rape	(4) Robbery	(5) Aggravated Assault
City Police Funds Per Resident	3.129	0.160**	-0.339	2.108*	-5.785
	(2.932)	(0.0764)	(0.259)	(1.105)	(8.788)
% of City Funds Spent on Police	5,216	76.64	296.0	674.4	5,780
	(3,559)	(92.49)	(289.6)	(1,230)	(9,776)
Population	0.00634***	5.22e-05***	0.000403***	0.00178***	0.00441***
	(0.000343)	(8.93e-06)	(3.03e-05)	(0.000129)	(0.00103)
Median Income	-0.0238	-0.000923	0.00129	-0.0144	0.0181
	(0.0450)	(0.00116)	(0.00399)	(0.0167)	(0.133)
Poverty Rate	17,371	341.0	1,207	1,528	20,774
	(13,886)	(359.6)	(1,228)	(5,203)	(41,368)
% High School Graduates	3,069	178.7	-320.8	-6,047	-2,630
	(12,178)	(316.7)	(1,060)	(4,505)	(35,820)
% Bachelor's Degree Holders	2,155	-9.368	-176.1	3,756*	-8,396
	(5,555)	(144.9)	(490.4)	(2,092)	(16,632)
Constant	-7,006	-198.9	272.4	3,720	2,510
	(12,699)	(327.8)	(1,108)	(4,674)	(37,164)
Observations	67	68	68	69	69
R-squared	0.883	0.511	0.792	0.826	0.295

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

7 Regressors with the hat represent the original endogenous variables which we used instrumental variables for.

Table 3: Property Crime Generic OLS Equation Results

VARIABLES	(1) Property Crime	(2) Burglary	(3) Larceny Theft	(4) Motor Vehicle Theft
City Police Funds Per Resident	-17.39	5.097	-5.963	1.761
	(12.48)	(3.852)	(8.571)	(3.020)
% City Funds Spent on Police	21,347	4,147	15,455	4,541
	(13,893)	(4,285)	(9,535)	(3,389)
Population	0.0173***	0.00162***	0.0140***	0.00121***
	(0.00146)	(0.000450)	(0.00100)	(0.000354)
Median Income	0.0761	-0.0351	-0.00373	-0.00739
	(0.189)	(0.0584)	(0.130)	(0.0460)
Poverty Rate	50,594	-8,790	7,763	-2,905
	(58,713)	(18,134)	(40,349)	(14,333)
% of High School Graduates	-92,933*	-18,754	-58,111	-23,827*
	(50,810)	(15,702)	(34,938)	(12,483)
% Bachelor's Degree Holders	53,153**	4,382	40,988**	7,174
	(23,682)	(7,291)	(16,222)	(5,836)
Constant	60,246	17,814	38,574	19,683
	(52,665)	(16,291)	(36,248)	(12,956)
Observations	68	69	69	67
R-squared	0.778	0.322	0.819	0.357

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

After running these original OLS regressions, we proceeded to model the relationship between our variables in a logarithmic model, and the results are depicted in tables 4, 5, and 6. As mentioned before, the crime statistics, "police funds per city resident," "population," and "median income" were all logged to make comparable relationships based on elasticities for these models in addition to finding a model of better fit. For these results, the population remained a statistically significant positive number across all crime statistics which is expected. This regression found a 1% increase in population to be correlated with about a 1% increase in crime across all crime statistics. In addition, an increase in police funding spent per city resident was correlated with a statistically significant increase in violent crime, murder, and robbery. Furthermore, a one percent increase in the percent of city funds spent on police correlated with a fairly substantial and statistically significant increase in property crime, burglary, and motor vehicle theft at 0.90%, 1.46%, and 1.70% respectively. This positive correlation between police spending and crime observed in the regression exhibits the bidirectional relationship between crime and police spending. The positive relationship may not be a result of the ineffectiveness of policing, but rather the fact that high crime areas are likely to spend more on law enforcement. This relationship highlights a severe shortcoming of the model as it does not account for the potential endogeneity. Therefore, while consistent with the literature and a substantial improvement to the metrics related to the goodness of fit relative to the OLS model, this logarithmic model is not able to provide us with definitive conclusions about the relationship at hand.

Table 4: Violent Crime Semi-Log Model Results

VARIABLES	(1) *log(Violent Crime)*	(2) *log(Aggravated Assault)*
log(City Police Funds Per Resident)	0.484*	0.303
	(0.280)	(0.305)
% City Funds Spent on Police	1.302*	1.373*
	(0.769)	(0.766)
log(Population)	1.075***	1.100***
	(0.0963)	(0.103)
log(Median Income)	-1.053	-1.593*
	(0.814)	(0.853)
Poverty Rate	0.308	0.603
	(3.577)	(3.804)
% High School Graduates	-0.0112	0.942
	(2.742)	(2.938)
% Bachelor's Degree Holders	1.336	1.499
	(1.302)	(1.403)
Constant	1.681	6.982
	(9.951)	(10.26)
Observations	67	69
R-squared	0.762	0.728

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Property Crime Cemi-Long Model Results

VARIABLES	(1) *log(Property Crime)*	(2) *log(Burglary)*	(3) *log(Larceny Theft)*
log(City Police Funds Per Resident)	-0.147	0.280	-0.139
	(0.213)	(0.287)	(0.210)
% City Funds Spent on Police	0.897*	1.456**	0.839
	(0.536)	(0.719)	(0.527)
log(Population)	0.991***	1.000***	0.973***
	(0.0722)	(0.0971)	(0.0711)
log(Median Income)	-0.900	-1.549*	-0.992*
	(0.599)	(0.800)	(0.586)
Poverty Rate	0.237	-2.425	-0.747
	(2.673)	(3.571)	(2.615)
% High School Graduates	-0.434	-0.529	-0.388
	(2.051)	(2.758)	(2.020)
% Bachelor's Degree Holders	2.793***	1.840	3.146***
	(0.980)	(1.317)	(0.965)
Constant	6.440	9.740	7.290
	(7.211)	(9.638)	(7.057)
Observations	68	69	69
R-squared	0.815	0.715	0.811

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Semi-Log Model Results with Robust Standard Erros (Murder, Rape, Robbery, and Motor-Vechile Theft)

VARIABLES	(1) log(Murder)	(2) log(Rape)	(3) log(Robbery)	(4) log(Motor Vehicle Theft)
log(City Police Funds Per Resident)	0.854**	0.0556	0.853**	0.496
	(0.361)	(0.203)	(0.341)	(0.436)
% City Funds Spent on Police	0.781	0.615	0.411	1.696**
	(0.853)	(0.516)	(0.564)	(0.778)
log(Population)	1.053***	0.972***	1.204***	1.048***
	(0.105)	(0.0657)	(0.104)	(0.137)
log(Median Income)	-2.356**	-1.219	-1.499	-1.633
	(1.084)	(1.035)	(0.977)	(1.043)
Poverty Rate	-0.101	-0.887	-1.038	-4.442
	(3.912)	(3.104)	(4.260)	(5.108)
% High School Graduates	0.485	1.152	-3.910	-0.575
	(3.423)	(2.969)	(2.825)	(2.603)
% Bachelor's Degree Holders	1.279	0.967	2.803*	1.805
	(1.958)	(1.915)	(1.582)	(1.667)
Constant	9.595	4.574	4.663	8.945
	(14.04)	(12.95)	(11.71)	(11.13)
Observations	68	68	69	67
R-squared	0.707	0.734	0.820	0.716

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Tables 7 and 8 below display the results of the 2SLS regression using the state tax rate and per capita state tax revenues as instruments for the log of city police funds per resident and the percentage of city funds spent on police. The coefficient of the population regressor represents a 1% increase in population correlated with about a 1% increase in crime and remains statistically significant except in the case of murder and violent crime. Beyond that relationship which was expected, our regression found no explanatory variable that was statistically significant consistently across the different types of crime. This finding is not inconsistent with the literature examined above. Numerous scholars have attempted to examine this relationship with many failing to establish a relationship between crime and policing. Additionally, limitations in data for police spending and crime statistics may contribute to the observations found in the model. This finding points to the difficulty in modeling complex issues such as crime and police spending. Our models did not find clear relationships between police spending and crime, given the broad range of the entire United States that this sample examines.

Table 7: Violent Crime 2SLS Regression Results[8]

VARIABLES	(1) log(Violent Crime)	(2) log(Murder)	(3) log(Rape)	(4) log(Robbery)	(5) log(Aggravated Assault)
log(City Police Funds Per Resident)*	3.835	-3.781	-2.486	-0.308	-2.896
	(77.14)	(16.57)	(2.722)	(2.043)	(3.299)
% City Funds Spent on Police*	31.05	0.890	1.627	2.680	6.937
	(380.9)	(82.81)	(9.516)	(6.617)	(10.69)
log(Population)	0.794	1.048	0.958***	1.191***	1.071***
	(3.521)	(0.896)	(0.145)	(0.115)	(0.186)
log(Median Income)	-1.556	2.287	1.262	-0.135	2.066
	(37.37)	(10.66)	(2.714)	(2.208)	(3.566)
Poverty Rate	-14.37	30.00	15.34	7.205	22.98
	(397.9)	(94.73)	(16.87)	(13.50)	(21.80)
% High School Graduates	6.380	-1.427	-0.131	-4.279	-0.162
	(102.9)	(18.91)	(4.487)	(3.277)	(5.293)
% Bachelor's Degree Holders	1.339	2.542	1.859	3.309**	2.831
	(12.14)	(3.357)	(2.116)	(1.653)	(2.670)
Constant	-14.70	-17.43	-9.250	-4.629	-17.32
	(76.34)	(36.35)	(22.21)	(17.11)	(27.63)
Observations	67	68	68	69	69
R-squared			0.284	0.767	0.190

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Property Crime 2SLS Regression Results

VARIABLES	(1) log(Property Crime)	(2) log(Burglary)	(3) log(Larceny Theft)	(4) log(Motor Vehicle Theft)
log(City Police Funds Per Resident)*	-2.445	-1.238	-2.982	-1.419
	(2.335)	(2.229)	(2.715)	(2.739)
% of City Funds Spent on Police*	4.078	-0.0443	5.346	-3.193
	(7.607)	(7.220)	(8.793)	(7.824)
log(Population)	0.974***	1.009***	0.948***	1.074***
	(0.131)	(0.126)	(0.153)	(0.154)
log(Median Income)	1.623	-0.329	2.261	-0.412
	(2.520)	(2.409)	(2.934)	(3.001)
Poverty Rate	15.90	6.435	19.13	5.563
	(15.41)	(14.73)	(17.94)	(18.26)
% of High School Graduates	-1.270	-1.473	-1.328	-2.131
	(3.727)	(3.576)	(4.355)	(4.480)
% Bachelor's Degree Holders	3.663*	2.111	4.295*	2.030
	(1.881)	(1.804)	(2.197)	(2.217)
Constant	-9.713	4.938	-14.39	7.003
	(19.45)	(18.67)	(22.74)	(23.22)
Observations	68	69	69	67
R-squared	0.438	0.560	0.194	0.415

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In regards to the 2SLS Regression, another potential limitation is the sample size we utilized for our data. Given the recency of our information, we only have police spending data on 70 major US cities and 2SLS may be unable to provide completely accurate and reliable estimates in small sample sizes (Jung, 2013). However, the 2SLS model is the most practical solution to mitigate the endogeneity of the police spending variables

8 Regressors with the asterisk represent the originally endogenous variables which we used instrumental variables for. Another note is that there is no unique definition of an R squared value and STATA uses different formulas than OLS in their computation.

is commonly used as a solution to this issue in previous studies. The problem of endogeneity is unable to be managed by simply utilizing the log model, therefore this model is the most useful option in this scenario.

5 Conclusion

Like any complicated social science study, there are a variety of limitations to our data and results. As one can see the fiscal year of 2020 (October 1, 2019 - September 30, 2020) and the 2019 calendar year do not line up perfectly. Unfortunately, at the time that this report was initiated the FBI only had a fully completed set of data for 2019 even though the data from 2020 clearly would have been preferred. However, there are still 3 overlapping months of data and it is unlikely that the average spending in these 70 cities shifted dramatically from 2019 to 2020. An additional limitation a study found is that many state and local governments spend more on correctional facilities than the raw police budgets, demonstrating that police spending is a small portion of the larger criminal justice system (Urban Institute, 2022). Furthermore, many other studies in the prior literature analyzed spending across time, which is more comparable to different practices in policing between different cities, and used race as a demographic variable which we did not consider[9]. City expenditures on policing may vastly differ in their results leading to ambiguous results across the entire nation.

For future studies, researchers could expand upon this work in a variety of different ways. When the full set of 2020 FBI statistics is released to the public, a comparison between these two years would yield informative results. The Vera study where we got our data on police spending is a unique piece of research that has not been well-used and may lead to many other findings. In addition, using a macro-level time-series approach across all of these cities as well as a more in-depth study about specific types of policing practices would be useful as well for anyone studying the impact of policing. Applying a time-series approach to a certain locality may be very helpful in examining this relationship. Police Departments in the United States are not monolitjthic institutions, their methods, spending, and approaches all vastly differ based on the locale. Crime as well is contingent on local conditions and factors which present opportunities for crime or motivate citizens to engage in illegal activity. Therefore, by analyzing demographic factors and police spending changes over the years in a single location, this relationship may be more clear

9 This report fouuses strictly on cross-sectional data

given that the study is repeated across many locales. The broad scope of this research which aims to cover the entire United States may be a limitation as local factors such as the state or city's legal frameworks and policing legislature are omitted. These factors may all affect the effectiveness of policing as a tool to reduce crime and point to the difficulty in examining this issue.

In conclusion, our analysis does not definitively demonstrate that an increase in police spending reduces crime. There are no statistically significant coefficients related to police expenditure found in our regression, which demonstrates that urban cities in our data that spent more money on policing did not have significantly smaller crime rates in 2019. One of the only consistent findings from our research is that population is positively correlated with crime rates. The major policy implications of these results include that while police spending may have ambiguous effects on crime across the nation, it is often a common solution that is utilized in reaction to crime. Our data does not show whether this is an effective solution or not, which is a much more complicated conclusion to make. Crime is a social phenomenon dependent on economic conditions, education, prevailing institutions, geography, and a myriad of other factors and one report will never be able to contain enough information to truly understand the solution to policing.

6 References

AD. (2016, August 7). Robust Standard Errors. *Economic Theory Blog.* https://economictheoryblog.com/2016/08/07/robust-standard-errors/

Annual Reports (2019) | *Chicago Police Department.* (n.d.). Retrieved March 28, 2022, from https://home.chicagopolice.org/statistics-data/statistical-reports/annual-reports/

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization, 6(3), 285–327.* https://doi.org/10.1177/1476127008094339

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy, 76(2), 169–217.* http://www.jstor.org/stable/1830482

Brier, S. S., & Fienberg, S. E. (1980). Recent Econometric Modeling of Crime and Punishment: Support for the Deterrence Hypothesis? *Evaluation Review, 4(2), 147–191.* https://doi.org/10.1177/0193841X8000400201

Bump, Philip. *Over the past 60 years, more spending on police hasn't necessarily meant less crime—The Washington Post.* (n.d.). Retrieved May 14, 2022, from https://www.washingtonpost.com/politics/2020/06/07/over-past-60-years-more-spending-police-hasnt-necessarily-meant-less-crime/

Cameron, S. (1994). A review of the econometric evidence on the effects of capital punishment. *The Journal of Socio-Economics, 23(1), 197–214.* https://doi.org/10.1016/1053-5357(94)90027-2

Crime Maps & Stats (2019) | Philadelphia Police Department. (n.d.). Retrieved March 28, 2022, from https://www.phillypolice.com/crime-maps-stats/

Crime Trends in California. (n.d.). Public Policy Institute of California. Retrieved March 4, 2022, from https://www.ppic.org/publication/crime-trends-in-california/

Criminal Justice Expenditures: Police, Corrections, and Courts. (2017, October 20). Urban Institute. https://www.urban.org/policy-centers/cross-center-initiatives/state-and-local-finance-initiative/state-and-local-backgrounders/criminal-justice-police-corrections-courts-expenditures

Cuomo, A. M., & Green, M. C. (2020). *New York State Crime Repor*t. 15.

Ehrlich, I. (1996). Crime, Punishment, and the Market for Offenses. *Journal of Economic Perspectives, 10(*1), 43–67. https://doi.org/10.1257/jep.10.1.43

Di Tella, R., & Schargrodsky, E. (2004). Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack. A*merican Economic Review, 94(*1), 115–133. https://doi.org/10.1257/000282804322970733

Goodman, J. D. (2021, October 10). A Year After 'Defund,' Police Departments Get Their Money Back. *The New York Times.* https://www.nytimes.com/2021/10/10/us/dallas-police-defund.html

Hallin, M. (2014). Gauss–Markov Theorem in Statistics. In *Wiley StatsRef: Statistics Reference Online.* John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118445112.stat07536

IBM. (2022, February 28). *Two-Stage Least-Squares Regression.* Retrieved April 7, 2022, from https://www.ibm.com/docs/fi/spss-statistics/SaaS-?topic=regression-two-stage-least-squares

Introductory Econometrics Chapter 19: Heteroskedasticity. (n.d.). Retrieved May 16, 2022, from http://www3.wabash.edu/econometrics/EconometricsBook/chap19.htm

Jung, S. (2013). Structural equation modeling with small sample sizes using two-stage ridge least-squares estimation. *Behav Res* 45, 75–81. https://doi.org/10.3758/s13428-012-0206-0

Kelly, M. (2000). Inequality and Crime. *The Review of Economics and Statistics, 82*(4),530–539. https://doi.org/https://doi.org/10.1162/003465300559028

Militarization fails to enhance police safety or reduce crime but may harm police reputation. (n.d.). PNAS. Retrieved March 4, 2022, from https://www-pnas-org.ccl.idm.oclc.org/doi/abs/10.1073/pnas.1805161115

Lin, M.-J. (2009). More police, less crime: Evidence from US state data. *International Review of Law and Economics,* 29, 73–80. https://doi.org/10.1016/j.irle.2008.12.003

Levitt, S. D. (1997). Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime. The American Economic Review, 87(3), 270–290. http://www.jstor.org/stable/2951346

Levitt, S. D. (2004). Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not. *Journal of Economic Perspectives*, 18(1), 163–190. https://doi.org/10.1257/089533004773563485

Lochner, L., & Moretti, E. (2004). T*he American Economic Review,* 94(1), 155–189. https://doi.org/https://eml.berkeley.edu/~moretti/lm46.pdf

McCrary, J. (2002). Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment. *The American Economic Review, 92(*4), 1236–1243. http://www.jstor.org/stable/3083311

*New Jersey. (*n.d.). FBI. Retrieved March 28, 2022, from https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/tables/table-8/table-8-stat e-cuts/new-jersey.xls

Pogue, T. F. (1975). Effect of Police Expenditures on Crime Rates: Some Evidence. Public Finance Quarterly, 3(1), 14–44. https://doi.org/10.1177/109114217500300102

Rushin, S., & Michalski, R. (2020). Police Funding. Florida Law Review, 72(2), 277-330. http://www.floridalawreview.com/2020/police-funding/

Scheider, M. C., & Mansourian, J. (2012). The Relationship Between Economic Conditions, Policing, and Crime Trends. In D. L. Spence (Ed.), The Impact of the Economic Downturn on American Police Agencies (pp. 1–17). essay.

Table 6. (n.d.). FBI. Retrieved March 4, 2022, from https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/tables/table-6/table-6

Theft Prevention—An overview (pdf) | ScienceDirect Topics. (n.d.). Retrieved March 4, 2022,
from https://www.sciencedirect.com/topics/computer-science/theft-prevention/pdf

Theft Prevention—An overview (pdf) | ScienceDirect Topics. (n.d.). Retrieved March 4, 2022,
from https://www.sciencedirect.com/topics/computer-science/theft-prevention/pdf

2019 State Tax Revenue. (n.d.). Retrieved May 14, 2022, from https://www.taxadmin.org/2019-state-tax-revenue

What Policing Costs: A Look at Spending in America's Biggest Cities. (n.d.). Vera Institute of Justice. Retrieved March 4, 2022, from https://www.vera.org/publications/what-policing-costs-in-americas-biggest-cities

Zerkel Mary. (2021, December 9). *6 reasons why it's time to defund the police.* American Friends Service Committee. Retrieved March 3, 2022, from https://www.afsc.org/blogs/news-and-commentary/6-reasons-why-its-time-to-defund-police

7 Appendices

Appendix: B-P and White Test Statistics for Generic OLS Equations[10]

VARIABLES	(1) Breusch-Pagan Test Statistic	(2) White Test Statistic
Violent Crime*	19.87	47.0474
Murder*	28.06	61.3156
Rape*	26.34	43.806
Robbery*	19.82	49.6593
Aggravated Assault	2.67	22.1835
Property Crime*	22.53	41.616
Burglary*	15.84	59.499
Larceny Theft*	13.05	39.1023
Motor Vehicle Theft*	21.41	59.0136

Appendix B: B-P and White Test Statstics for Semi-Log Model

VARIABLES	(1) Breusch-Pagan Test Statistic	(2) White Test Statistic
Violent Crime	5.5679	46.7794
Murder*	12.0904	44.0096
Rape*	11.429	55.8484
Robbery*	12.869	36.8115
Aggravated Assault	7.8246	37.6119
Property Crime	8.506	41.412
Burglary	8.28	38.4606
Larceny Theft	9.9498	44.498
Motor Vehicle Theft*	12.8159	46.7928

10 Asterisks denote that the null hypothesis of no heteroskedasticity was rejected for both tests.

Does High-Speed Internet Impact Knowledge Diffusion? Evidence from Patent Citations

Yunshu Hu

May 2022

Abstract

Motivated by the Information Communications Technology productivity paradox and the lack of renewal in theories regarding knowledge diffusion, this paper examines the impact of broadband, or high-speed internet, on knowledge exchange, traced by patent citations, using publicly available county-level data from USPTO and CPS. Both fixed effect and instrumental variable approaches are employed to reduce endogeneity concerns. While gaining access to broadband can lead to counties that do not cite each other having citations in between for the first time, a higher internet penetration rate shows an ambiguous impact on increased knowledge exchange.

A thesis submitted for the degree of Bachelor of Arts in Economics with Honors at the University of Chicago

Acknowledgement

I would like to express my gratitude to my advisor, Christopher Dobronyi, who guided me throughout this project. I wish to acknowledge the help and insights provided by the faculty at Kenneth C. Griffin Department of Economics at the University of Chicago, including but not limited to Victor Lima, Kotaro Yoshida, and Alexander Torgovitsky. I would also like to thank Yuerong Zhuang, Eazin Pu, and my mom for their emotional support.

Content

1 Introduction

Since its popularization in the 1990s, broadband, or a permanently connected high-speed internet line, is believed to be crucial to the economic growth and the diffusion of knowledge. However, the rapid development of communication technologies over the past 20 years has coincided with a generalized slowdown in aggregate productivity growth, which is called the modern productivity paradox (Acemoglu et al., 2013). For instance, while many firms now have access to broadband networks, the diffusion of more advanced digital tools and applications is far from complete and differs significantly across countries (McKinsey Global Institute, 2018). Therefore, the question of whether high-speed internet plays a role in knowledge diffusion becomes intriguing.

On the other hand, it is also not easy to quantify the extent and impact of knowledge spillovers. "Knowledge flows are invisible; they leave no paper trail by which they may be measured and tracked, and there is nothing to prevent the theorist from assuming anything about them that she likes." (Krugman, 1993) Social science researchers have sought for various ways to proxy knowledge diffusion in the past decades. In this article, I plan to analyze the diffusion of knowledge using one of the common approaches: utilize patent citations to identify a "paper trail" that may be associated with knowledge flows between geographical locations.

Abundant literature regarding determinants of patent spillover accepts the conclusion of Jaffe, Trajtenberg, and Henderson (1993) that knowledge spillovers are localized, and uses the dataset the authors have constructed - patent citations from 1975 to 1999. For example, MacGarvie (2005) examined the determinants of international patent diffusion using the JTH dataset. The time frame of the JTH data limits its capability of incorporating the influence of high-speed internet, so this article will fill this blank. I employ the USPTO patent citation data and the CPS broadband data to study the impact of broadband availability on number of citations on county level from 2001 to 2008, when broadband coverage in the United States changed dramatically. Constructed a matrix in which between each pair of counties, I studied whether there is a causal effect of internet availability of both counties on the number of citations between them using various model specifications. A gravity model is innovatively employed to study the impact on citation volume. The results suggest that gaining access to broadband can bring about citation activities between two counties that previously do not have citations in between, but there is no evidence on an impact of broadband on magnitude.

Section 2 of this paper is a more detailed literature review. Section 3 will lay out the theoretical models; Section 4 will elaborate the sources of data; Section 5 will analyze the regression results, and Section 6 will be conclusions using the current results.

2 Literature Review

The literature review on the topic of how the Internet impacted knowledge diffusion is done from two directions. On the one hand, existing literature has made remarkable efforts in exploring the determinants of knowledge diffusion; on the other hand, the uneven rollout of broadband internet has been a commonly studied independent variable in industrial or ganization research. By examining what determines knowledge diffusion and how access to broadband internet can influence or modify these determinants, we can remodel knowledge diffusion after the installation of broadband internet.

2.1 The Determinants of Knowledge Diffusion

It is commonly accepted that knowledge spillovers are highly localized, especially when using patent citation data (Jaffe et al. 1993, hereafter JTH). Despite Krugman's comment that "there is nothing to prevent the theorist from assuming anything about [knowledge flows] that she likes", interpreting patent citations as a measure of knowledge flows is validated by various research (Duguet and MacGarvie 2005). Thus, in this paper, I will continue using patent citations to trace knowledge diffusion.

Studies on the determinants of patent citations have reached a consensus that geography is one of the main factors in the diffusion of ideas. (Carlino and Kerr 2014) The seminal study of JTH, making use of the detailed location information of patents in 1975 and 1980, shows a "home bias" in patent citations, i.e. inventors are more likely to cite other inventors who are geographically close to them. This finding is extended by different researchers. Murata et al. (2014) used a distance-based approach to strengthen the argument that excluding the impact of state or metropolitan borders, distance is a significant determinant of patent citations.

The geographic concentration of research institutes, universities, as well as corporate R&D labs, are strongly correlated with the agglomeration of patent citations. Although lacking causal inference, the human capital density in US cities is the most important correlate of patenting rates per capita. (Carlino et al. 2007) City size can also be a determinant of knowledge exchange. Berliant et al. (2006) used a random-matching model to theoretically hypothesize that a higher population mass leads to more frequent innovative activity. The studies over international knowledge diffusion using the original JTH dataset indicate that foreign direct investment, common language, and technology similarity are additional significant factors. (MacGarvie 2005) Knowledge can also transfer through multinational firms, though the spatial barriers still exist, and the firm revenue will decrease in the intensity of the knowledge transfer. (Keller and Yeaple 2009)

Numerous pieces of literature echo the conclusion that distance matters to knowledge exchange. Yet there is a presumption that distance has been becoming less important over time with improvement in communication and transport links, which provokes this paper. I will test this hypothesis by first looking at the other side of the story – the development of broadband, or high-speed Internet.

2.2 The Industrial-Organizational Impact of Broadband Internet

Since the invention of computers in the middle of the last century, what can be broadlylabeled as Information Communications Technology (ICT) has become faster, cheaper, more important, and ubiquitous. Personal computers became available for individuals and small businesses in the 1980s, soon giving rise to the need for networked PCs. The invention of DNS, the common use of TCP/IP, and the popularity of email caused an explosion of activity on the Internet. Between 1986 and 1987, the network grew from 2,000 hosts to 30,000. In the meantime, telecommunications providers have continued to innovate and invest in improving the "bandwidth" of the network, which has permitted ever-increasing speeds of communication.

According to the Federal Communications Commission (FCC), broadband, or high-speed internet refers to advanced telecommunication services with over 200 kbps capability in at least one direction. Since the late 1990s, broadband infrastructure developed rapidly in the United States and worldwide. In light of modern theories of endogenous growth (Romer 1990), broadband infrastructure allows for the generation and dissemination of decentralzed knowledge and could accelerate economic

growth by facilitating the development and adoption of innovation processes.

The Information Communications Technology (ICT) productivity paradox used to be an issue; works by Morrison and Berndt (1991), Strassman (1990), and Roach (1987), for example, failed to find measurable benefits attributable to ICT. Robert Solow famously quipped that "we see computers everywhere but in the productivity statistics." (1987) In hindsight after 20 years, a common explanation of the productivity paradox was that ICT, including broadband, is a general-purpose technology that is used by businesses in many ways to produce various intermediate and final goods and services. Many of the benefits of broadband usage stem from its enabling capability, which allows for the development of new business processes and work practices.(Bresnahan and Trajtenberg 1995) Some other explanations of the lagged effect of ICT investments include: it takes time for firms to absorb the organizational cost of adopting new technologies (Torres et al. 2010); the then prevalent accounting measures and price indices needed to be modified to capture the value of intangible assets (Brynjolfsson and Hitt 2000), etc. Using data from a longer time span, recent literature suggests that broadband installation contributed to the growth in average earnings (Mack and Faggian 2013) and overall GDP (Czernich et al. 2010)

However, the impact of broadband infrastructure on knowledge diffusion remains to be obscure. The causal relationship between the spatial distribution of broadband and the knowledge clusters of firms and R&D institutions is heterogeneous. Intuitively, broadband appears to be an enabling technology that allows firms and inventors to strategically locate in lower-cost counties and closer to major knowledge centers in the United States. But whether the communications via broadband Internet connections can overcome the negative externalities associated with knowledge agglomeration is yet to be discovered. (Mack 2012) This paper aims to revisit this question using patent citation data and the change in broadband accessibility in United States counties.

3 Methods

There are two interesting dependent variables regarding the impact of broadband on United States patent citations: whether there exists a patent citation between two counties and how many citations there are between two counties.

The potential issue of studying the impact of high-speed internet lies in endogeneity. The diffusion of broadband is not random, since this technology is primarily provided by private sector firms, and the firms build the infrastructure in places where it is more prof itable. Such places may show technology agglomeration. In addition, there also can be reverse causality, where locations with more innovation and R&D tends to attract broad band providers. (Atasoy, 2013) I will address the endogeneity issue in several ways, including fixed effects and instrumental variables.

3.1 From Zero to One: Citation Existence

The first dependent variable of interest is whether citation exists between two counties. Assuming the conditional choice probability of at least a citing activity existing between two counties is of a standard logistic distribution, I started with a simple binary choice logit model:

$$\mathbb{P}[Citation_{ijt} = 1|\boldsymbol{X_{ijt}}] = \frac{1}{1+\exp(-\boldsymbol{X'_{ijt}\beta})} \tag{1}$$

in which the independent variables include

$$\begin{aligned} \boldsymbol{X}_{ijt} =& \{BroadbandAvailability_{it}, BroadbandAvailability_{jt}, \\ & NumCitingCountyPatents_{it}, NumCitingCountyPatents_{jt}, \\ & Distance_{ij}, Population_{it}, Income_{it}, \\ & Population_{jt}, Income_{jt}, Year_t\} \end{aligned}$$

The major coefficients of interest are the broadband availability in counties *i* and *j*, where county *i* is the citing county and county *j* is the cited county. The other control variables are chosen based on previous literature review that distance, region size, and existing R&D are the major determinants of patent citation. I argue that the number of patents granted in each county each year is sufficient to proxy the local R&D level, and is more time-variant than the number of R&D institutions or firms. The amount of R&D investments is often considered as an ideal variable to reflect the existing technological and innovational level of a region, but due to data limitation, I decided to use the number of patents granted only. The size of the county is captured by the population and income of that year. If the naive assumption on the conditional distribution of the dependent variable is true, we can obtain the direction of the independent variables' impacts on the citation activity between two counties.

With no evidence that the assumption holds, I additionally included a fixed effects regression to increase the robustness.

$$\begin{aligned} Citation_{ij} =& \beta_1 BroadbandAvailability_{it} + \beta_2 BroadbandAvailability_{jt} + \\ & \gamma_1 NumCitingCountyPatents_{it} + \gamma 2 NumCitedCountyPatents_{jt} + \\ & \gamma_3 Distance_{ij} + a_i + v_t + u_{it} \end{aligned} \tag{2}$$

The fixed effects model assumes that the regional-specific effect (such as concentration of R&D institutions) on whether a citation occurs between two counties are fixed is time invariant; the year-specific effect is region-invariant. The assumptions are reasonable since the clustering effect of patent citation is enduring over time and all patents are granted through the national organization United States Patent and Trademark Office without special considerations on regions. (Carlino et al. 2012)

Both the logit model and certain fixed effect specifications suggest that installing high speed Internet increases the probability of existing citation between two counties. This will be analyzed in the results section.

3.2 The Gravity Model: Citation Volume

3.2.1 Applying Gravity Model to Knowledge "Trade"

This paper innovatively utilizes the gravity model in international trade to capture the citation volume between United States Counties. The gravity model has gained success in explaining bilateral trade flows by describing the trade volume between two economies as increasing in their economic size and decreasing in their distance. The model has been applied to relations other than trade in goods such as services (Ceglowski 2006). Notably, Peri (2005) applied the gravity framework to analyze international patent citations in 1975 - 1996, showing that the localization in patent citation is significant, though slightly decreasing in the 10-year time span. Blum and Goldfarb (2005) employed the gravity model to alyze the website visits of the United States Internet users in 2000 and found that the negative distance effect is preeminent on digital products related to taste but not on software. Picci (2010) examined patent collaboration using the gravity model, and concluded that there exists a negative distance effect but significantly smaller than the mean elasticity of 0.9 in international trade.

I assume that the number of patent citations between two counties i and j at time t follows the relationship

$$NumCitations_{ijt} = \frac{X_{ijt}(NumCitingCountyPatents_{it})^{\beta_1}(NumCitedCountyPatents_{jt})^{\beta_2}}{(Distance_{ij})^{\gamma}}$$

where X_{ijt} are county and time specific factors. The equation suggests that the number of citations between two counties is proportional to the existing patents in both counties and inversely proportional to the distance between the two counties. This is consistent with the previous finding that knowledge diffusion increases in the concentration of local R&D level and has a localization effect. Taking logs of both sides gives the basic regression equation

$$\begin{aligned}\log(NumCitations)_{ijt} = &\log(X_{ijt}) + \beta_1 \log(NumCitingCountyPatents)_{it}\\ &+ \beta_2 \log(NumCitedCountyPatents)_{jt})\\ &+ \gamma Distance_{ij} + v_t + u_{ijt}\end{aligned} \tag{3}$$

The feasibility of this basic model is verified using patents citations from 1996 to 2008 when assuming X is a constant.

3.2.2 Impact of Broadband: Fixed Effects Approach

If X is not a constant, then it may be proportional to a group of covariates, including the availability of broadband, as well as the region size. To study the impact of high-speed Internet, the model is extended into the following

$$\begin{aligned}\log(NumCitations)_{ijt} = &\alpha_1 Broadband_{it} + \alpha_2 Broadband_{jt}\\ &+ \beta_1 \log(NumCitingCountyPatents)_{it}\\ &+ \beta_2 \log(NumCitedCountyPatents)_{jt})\\ &+ \gamma_1 Distance_{ij} + \gamma_2 Population_{it} + \gamma_3 Income_{it}\\ &+ \gamma_4 Population_{jt} + \gamma_5 Income_{jt} + v_t + u_{ijt}\end{aligned} \tag{4}$$

Where broadband can be either an indicator variable of broadband installation or a continuous variable of log Internet penetration rate. We can use the fixed effect regressions similar to equation (2), with the assumption that the unobserved region-specific and time-specific effects

on the citation volume can be captured by different intercepts.

3.2.3 Impact of Broadband: Instrumental Variable Approach

It is reasonable to doubt whether the fixed effects model can solve the reverse causality concern. A region with more research and innovation may induce more telecommunications providers to compete and lower the price for Internet infrastructure, thus increasing the penetration rate.

I will address the concern of endogeneity by an instrumental variable and two stage least squares approach, although the selection of instrumental variable is open to further discussion. I currently selected two different IVs: one is the average broadband availability of adjacent counties, the other is the terrain ruggedness.

For each county, I calculate the average internet penetration rate of all of its adjacent counties as an instrument. The assumption is that the instruments are correlated with broadband penetration in a target county but remain uncorrelated with citation outcomes. The relevance assumption is supported by regression results, where the first stage regression gives a significant 1.02 coefficient. Although there is still concern that the clustering effect of innovation can be preeminent in areas larger than a single county, county adjacency is not limited the political border of states, thus the mean internet speed can better represent the exogenous broadband infrastructure of the area. This approach is also seen in a study on the impact of broadband penetration on U.S. farm productivity (LoPiccalo, 2020).

The first stage of this IV is given below

$$BBAvailability_{it} = \alpha_1 + \gamma_1 AvgAdjacentBBAvailability)_{it} + \boldsymbol{C_{ijt}\gamma_2} + m_t + \epsilon_{ijt}$$

$$\log(BBPenetration)_{it} = \alpha_2 + \gamma_3 \log(AvgAdjacentBBPenetration)_{it} + \boldsymbol{C_{ijt}\gamma_4} + n_t + \phi_{ijt}$$

where $\mathbf{C}_{ijt}$ is the same set of control variables in the fixed effects regression. I also used the terrain ruggedness index developed by Riley et al. (1999) to express the amount of elevation of a region as an instrumental variable for a county's broadband penetration. This approach is also adopted by the study of Kolko (2012) on the impact of broadband on economic growth. To be a good instrument, terrain ruggedness should be correlated with broadband availability without being independently

correlated with patent citation. Broadband providers face higher costs to extend service in areas with steeper terrain, which is confirmed in the first stage regression that gives a significant negative coefficient of interest. Remarkably, the terrain ruggedness index is not time-variant, so I multiply the terrain ruggedness index with a year indicator.
The first stage of this IV is given below.

$$BBAvailability_{it} = \alpha_1 + \gamma_1 Ruggedness_i \times Year_{it} + \boldsymbol{C_{ijt}\gamma_2} + m_t + \epsilon_{ijt}$$

$$\log(BBPenetration)_{it} = \alpha_2 + \gamma_3 Ruggedness_i \times Year_{it} + \boldsymbol{C_{ijt}\gamma_4} + n_t + \phi_{ijt}$$

Thus, the second stage of the 2SLS regressions are

$$\begin{aligned}\log(NumCitations)_{ijt} = {} & \beta_1 BBAva\hat{i}lability_{it} + \beta_2 BBAva\hat{i}lability_{jt} \\ & + \boldsymbol{C_{ijt}\beta_3} + v_t + u_{ijt}\end{aligned} \tag{5}$$

The results are discussed in section 5. While the gravity performs well, the impact of the high-speed internet on citation volume is ambiguous

4 Data

The data that I utilize mainly come from two sources: USPTO patent citation data from PatentsView, the broadband data and the county demographics from the Census. PatentsView is a data visualization, dissemination, and analysis platform that focuses on intellectual property (IP) data, supported by the Office of the Chief Economist at the U.S. Patent and Trademark Office (USPTO). From this platform, I obtained clean patent citation data where each patent is matched to the other patents that it is citing. I merged the citation dataset with the patent assignee dataset and location dataset, where I gained the county where the assignees of the citing and cited patents are located. I limited my sample to citations between U.S. patents in order to eliminate the potential unobserved confounders in international citation. I also eliminated self-citations, which does not capture knowledge flow.

Although the Federal Communications Commission provides broadband connection data, the information is provided by the broadband companies. This is also one of the main reason I did not choose internet speed data, which is reported as the maximum speed that the companies can provide, rather than the actual speed. I chose the CPS data that reports the internet connection from the households side. The Center on Technology, Data and Society at Arizona University provides county-level time series data that contains the yearly estimates of the Internet penetration rate in U.S. counties

I merged the patent citation data and the broadband availability data, and reconstructed a matrix in which each county is paired to each of the other counties once from year 2001 to 2008. I chose a long time lag to account for the potential lag effect of the ICT investment. I derived the number of citations between each pair of counties, and added the distance between the two counties. Then for each county, I added broadband variables: whether the county has access to high-speed internet and the broadband penetration rate. Also, I obtained county demographics data from the Census as the control variables. A summary statistics table is shown in Table 1.

Using the "marginal" counties that switched from no high-speed internet access to having high-speed internet access, I found that for each county, the number of citations with closer counties (the counties that are within the first quantile of distance in the sample) decreased and the number of citations with counties farther away are increased, as shown in Figure 1. This finding serves as the motivation for the regressions that follow.

5 Results

Several regressions are ran as described in section 3. The full regression results are displayed in the Appendix. I will examine each set of regressions in this section.

5.1 Citation Existence

Table 2 displays the regression results of the logit model in equation (1) and the fixed effects models regarding the impact of broadband affability on the existence of citation between two counties. The dependent variable is an indicator taking value 1 for citation activity existing between the county pair. Due to the fact that over 90% of the county pairs do not have patent citations between them in 2001 to 2008, the coefficients are small. The coefficients on year indicators in the logit model are not included.

The logit model has coefficients of interests 0.683 and 0.727, suggesting that broadband availability has a significant positive effect on the existence of citation. A county with connection to high-speed internet has e0.683 = 1.98 times the odds of a county without connection to high-speed internet of citing patents from outside. On the other hand, a county with connection to high-speed internet has e0.727 = 2.07 times the odds of a county without connection to high-speed internet of being cited.

Column (2) of Table 2 shows the year fixed effects regression result. It implies that if the citing county has broadband, then there will be a 0.007 additional chance for the existence of citing activity.

Column (3) of Table 2 shows the year plus citing county fixed effects. When adding the citing county fixed effect, only the broadband availability of the cited county has a significant impact on citation activity, indicating that when a county has connection to broadband, patents originated from here will be cited with an additional 0.008 probability. These results imply that gaining access to high-speed Internet leads to a higher chance of having citation activities.

5.2 Citation Volume

I first tested the feasibility of the gravity model described in equation (3) using the yearly citation data from 1996 to 2008. The results are shown in Table 3. No matter controlling for the county demographics or not, using OLS or year fixed effects, the gravity model performs well. Though there exists potential omitted variable, it confirms the previous literature on the determinants of patent citations. The gravity model suggests that there is a 0.008 - 0.009 % decrease in patent citations when the distance of two counties increases by 1%. One percent additional existing patent can lead to up to 0.01% increase in citation volume. Given these results, I proceeded by adding broadband variables as described in equation (4) and applying the model to the sample of 2001-2008 data.

The following table is a quick summary of the coefficients of interest in the rest of the regressions. The full results are listed in the Appendix.

	BB Exists (Citing cty)	BB Exists (Cited cty)	log(BB Speed) (Citing cty)	log(BB Speed) (Cited cty)
FE	0.002***	0.0001	-0.0004***	0.0002
IV1	0.004***	0.004***	-0.001***	-0.0001
IV2	0.029***	0.028***	-0.007*	0.003

The results of the year fixed effects model are shown in Table 4. Due to the fact that adding county fixed effects caused rank-deficient or indefinite problems, I used year fixed effects only. Column (1) uses the indicator variables of whether the county has access to broadband as the independent variable, and column (2) uses the log internet penetration rate. While the gravity model still performs well, there is statistical significance in the broadband availability and penetration of the citing county. Column (1) suggests that if the county has access to high-speed internet, there will be 0.002% increase in citing from other counties; column (2) suggests that if there is a 1% increase in Internet penetration rate, there will be 0.0004% decrease in citing from other counties. Both specifications suggests that the internet infrastructure at the citing county has a stronger impact on citation activities.

The results of the IV models are shown in Table 5. Column (1) and (2) are the results of the adjacent average broadband IV approach; column (3) and (4) use the terrain ruggedness IV approach. While the gravity model still performs well, column (1) suggests that if the county has access to high-speed internet, there will be 0.004% increase in both citing and being cited from other counties; column (2) suggests that if there is a 1% increase in Internet penetration rate, there will be 0.001% decrease in citing from other counties; column (3) indicates that if the county has access to high-speed internet, there will be 0.028-0.029% increase in both citing and being cited from other counties; column (4) indicates that if there is a 1% increase in Internet penetration rate, there will be 0.007% decrease in citing from other counties. The coefficients on the broadband indicator variables are exceptionally large for the ruggedness IV, comparing to the results of the other IV regression and the fixed effects regression. While this does not necessarily infer which IV is better, one explanation is that the terrain ruggedness is actually endogenous and highly correlated the citation volume of a region. Intuitively, many regions with more geographic elevations are rural areas, which might link to less citation activities. The remarkable difference in the coefficients of interest in columns (1) and (3) implies that more appropriate instrumental variables for broadband availability are yet to be found.

While the indicator variable of whether there is high-speed Internet access always plays a role in the gravity model, the broadband penetration rate seems to be not as significant. It is surprising that when a larger proportion of the population are using high-speed internet, the citing activity declines. This result in consistent in fixed effects and IV specifications, and requires further investigation.

6 Conclusion

This paper shows the feasibility of applying the gravity model to the research in patent citation and knowledge flow, and this model can be employed in future studies on the de terminants of knowledge diffusion. Using data from 2001 to 2008, I found that broadband infrastructure changing from zero to one may lead to citation between two counties changing from zero to one. Having high-speed internet access implies more knowledge exchange taking place. The increase in the penetration rate of broadband, however, cannot bring an increase in number of citations. This agrees with the productivity paradox in ICT – technology im provement cannot lead to output growth. On the other hand, the result might be limited by data: the CPS broadband data are imputed from households rather than R&D institutions, where a lot of innovation actually took place.

There are several potential ways to improve the robustness of this paper. First, the current models does not control for the technological similarity between each pair of counties. Younge and Kuhn (2020) constructed a vector space model that compares the text of every patent to every other patent granted by the USPTO. This measure allows us to determine the technological distance between each citing/cited pair of patents based on the technical description of each patent. The authors found that there is big variation among the textual similarity of the cited and citing patents. Adding a variable that captures the average technological similarity of all citations between two counties may improve the accuracy of the specification. Furthermore, there may be better instrumental variables to be employed. The current IVs do not fully eliminate the concern of innovation agglomeration.

This paper fills the blank in two-folds: the lack of consideration of internet in patent related research and the lack of exploration of knowledge diffusion in broadband-related research. More research needs to be done in this intersection: if internet does not influence knowledge diffusion, what factor in the 21st century does? Or if internet does not alter patent citations, what part of knowledge exchange does internet changed? It is surprising the the knowledge diffusion pattern today might still be the same as that of 50 years ago, after human beings have changed their ways of living dramatically.

7 References

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr. "Innovation, Reallocation and Growth," 2013. https://doi.org/10.3386/w18993.

Atasoy, Hilal. "The Effects of Broadband Internet Expansion on Labor Market Out comes." ILR Review 66, no. 2 (2013): 315–45. https://doi.org/10.1177/001979391306600202.

Berliant, Marcus, Robert R. Reed, and Ping Wang. "Knowledge Exchange, Matching, and Agglomeration." Journal of Urban Economics 60, no. 1 (2006): 69–95. https://doi. org/10.1016/j.jue.2006.01.004.

Blum, Bernardo S., and Avi Goldfarb. "Does the Internet Defy the Law of Gravity?" Journal of International Economics 70, no. 2 (2006): 384–405. https://doi.org/10.1016/ j.jinteco.2005.10.002.

Bresnahan, Timothy F., and M. Trajtenberg. "General Purpose Technologies 'Engines of Growth'?" Journal of Econometrics 65, no. 1 (1995): 83–108. https://doi.org/10. 1016/0304-4076(94)01598-t.

Brynjolfsson, Erik, and Lorin M Hitt. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." Journal of Economic Perspec tives 14, no. 4 (2000): 23–48. https://doi.org/10.1257/jep.14.4.23.

Carlino, Gerald A., Satyajit Chatterjee, and Robert M. Hunt. "Urban Density and the Rate of Invention." Journal of Urban Economics 61, no. 3 (2007): 389–419. https: //doi.org/10.1016/j.jue.2006.08.003.

Carlino, Gerald, and William Kerr. "Agglomeration and Innovation," 2014. https: //doi.org/10.3386/w20367.

Ceglowski, Janet. "Does Gravity Matter in a Service Economy?" Review of World Economics 142, no. 2 (2006): 307–29. https://doi.org/10.1007/s10290-006-0069-5.

Czernich, Nina, Oliver Falck, Tobias Kretschmer, and Ludger Woessmann. "Broadband Infrastructure and Economic Growth." The Economic Journal 121, no. 552 (2011): 505–32. https://doi.org/10.1111/j.1468-0297.2011.02420.x.

Duguet, Emmanuel, and Megan MacGarvie. "How Well Do Patent Citations Mea sure Flows of Technology? Evidence from French Innovation Surveys." Economics of In novation and New Technology 14, no. 5 (2005): 375–93. https://doi.org/10.1080/ 1043859042000307347.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," 1992. https://doi.org/10. 3386/w3993.

Keller, Wolfgang, and Stephen Yeaple. "The Gravity of Knowledge," 2009. https: //doi.org/10.3386/w15509.

Kolko, Jed. "Broadband and Local Growth." Journal of Urban Economics 71, no. 1 (2012): 100–113. https://doi.org/10.1016/j.jue.2011.07.004.

Kuhn, Jeffrey, Kenneth Younge, and Alan Marco. "Patent Citations Reexamined." The RAND Journal of Economics 51, no. 1 (2020): 109–32. https://doi.org/10.1111/ 1756-2171.12307.

LoPiccalo, Katherine. "Impact of Broadband Penetration on U.S. Farm Productivity." SSRN Electronic Journal, 2021. https://doi.org/10.2139/ ssrn.3790850.

MacGarvie, Megan. "The Determinants of International Knowledge Diffusion as Mea sured by Patent Citations." Economics Letters 87, no. 1 (2005): 121–26. https://doi. org/10.1016/j.econlet.2004.09.011.

Mack, Elizabeth A. "Broadband and Knowledge Intensive Firm Clusters: Essential Link or Auxiliary Connection?" Papers in Regional Science 93, no. 1 (2012): 3–29. https: //doi.org/10.1111/j.1435-5957.2012.00461.x.

Mack, Elizabeth, and Alessandra Faggian. "Productivity and Broadband." Interna tional Regional Science Review 36, no. 3 (2013): 392–423. https:// doi.org/10.1177/ 0160017612471191.

Martinez, Diego, Jesus Rodriguez, and Jose L. Torres. "ICT-Specific Technological Change and Productivity Growth in the US: 1980–2004." Information Economics and Policy 22, no. 2 (2010): 121–29. https://doi. org/10.1016/j.infoecopol.2009.07.001.

Morrison, Catherine, and Ernst Berndt. "Assessing the Productivity of Information Technology Equipment in U.S. Manufacturing Industries," 1991. https://doi.org/10.3386/w3582.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura. "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." Review of Eco-nomics and Statistics 96, no. 5 (2014): 967–85. https://doi.org/10.1162/rest_a_00422.

Peri, Giovanni. "Determinants of Knowledge Flows and Their Effect on Innovation." Review of Economics and Statistics 87, no. 2 (2005): 308–22. https://doi.org/10.1162/ 0034653053970258.

Picci, Lucio. "The Internationalization of Inventive Activity: A Gravity Model Using Patent Data." Research Policy 39, no. 8 (2010): 1070–81. https://doi.org/10.1016/j. respol.2010.05.007.

Riley, S.J., S.D. DeGloria and R. Elliot. "A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity". Intermountain Journal of Sciences 5(1-4) (1999):23-27 Roach, Stephen S. America's Technology Dilemma: A Profile of the Information Econ omy. New York: Morgan Stanley, 1987.

Roach, Stephen S. America's Technology Dilemma: A Profile of the Information Economy. New York: Morgan Stanley, 1987.

Strassmann, Paul A. The Business Value of Computers: An Executive's Guide. New Canaan, CT: Information Economics Press, 1991.

9 Appendix: Tables and Graphs

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
NumCitations	977,751	0.039	1.091	0	0	0	404
Citation	977,751	0.017	0.129	0	0	0	1
NumPatents_i	977,751	14.398	52.145	1	1	8	1,010
NumPatents_j	977,751	244.634	845.814	1	14	153	12,953
Distance	977,751	1,005.519	698.729	0	484.7	1,427.3	5,134
year	977,751	2,003.357	2.481	2,000	2,001	2,005	2,008
if_bb_i	977,751	0.384	0.486	0	0	1	1
if_bb_j	977,751	0.384	0.486	0	0	1	1
broadband_i	977,751	0.130	0.210	0	0	0.2	1
broadband_j	977,751	0.130	0.210	0	0	0.2	1

Figure 1: Counties That Changed From Having No Broadband to Having Broadband

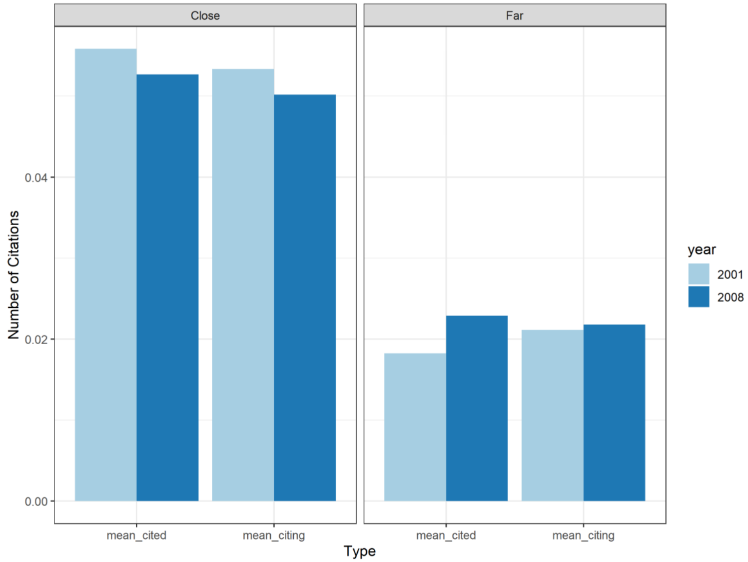


Table 2: Dependent Variable: Citation Existence

	Dependent variable:		
	Citation		
	logistic	*felm*	
	(1)	(2)	(3)
if_bb_i	0.683^{***}	0.007^{***}	0.002
	(0.017)	(0.002)	(0.001)
if_bb_j	0.727^{***}	0.007^{***}	0.008^{***}
	(0.017)	(0.001)	(0.0003)
Distance	-0.0002^{***}	-0.00000^{***}	-0.00001^{***}
	(0.00001)	(0.00000)	(0.00000)
NumPatents_i	0.005^{***}	0.0005^{***}	0.0001^{***}
	(0.0001)	(0.0001)	(0.00001)
NumPatents_j	0.0003^{***}	0.00003^{***}	0.00003^{***}
	(0.00000)	(0.00000)	(0.00000)
Constant	-4.885^{***}		
	(0.027)		
Observations	1,013,596	1,013,596	1,013,596
R^2		0.081	0.094
Adjusted R^2		0.081	0.094
Log Likelihood	−73,835.640		
Akaike Inf. Crit.	147,699.300		
Residual Std. Error		0.122 (df = 1013582)	0.121 (df = 1012855)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: Test of Gravity Model

	Dependent variable:			
	log_NumCitations			
	OLS	*felm*	*OLS*	*felm*
	(1)	(2)	(3)	(4)
log_NumPatents_i	0.010^{***}	0.010^{***}	0.010^{***}	0.010^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
log_NumPatents_j	0.006^{***}	0.006^{***}	0.005^{***}	0.005^{***}
	(0.00004)	(0.00004)	(0.0001)	(0.0001)
log_Distance	-0.008^{***}	-0.008^{***}	-0.008^{***}	-0.008^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
log(pop_i)			-0.001^{***}	-0.001^{***}
			(0.0001)	(0.0001)
log(pop_j)			0.0001	0.0001
			(0.0001)	(0.0001)
log(income_i)			-0.002^{***}	−0.0001
			(0.0003)	(0.0004)
log(income_j)			0.004^{***}	0.007^{***}
			(0.0003)	(0.0004)
Constant	0.025^{***}		0.018^{***}	
	(0.001)		(0.004)	
Observations	1,591,872	1,591,872	1,523,272	1,523,272
R^2	0.035	0.035	0.037	0.037
Adjusted R^2	0.035	0.035	0.037	0.037
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4: Dependent Variable: Citation Volume (FE Regressions)

	Dependent variable:	
	log_NumCitations	
	(1)	(2)
if_bb_i	0.002***	
	(0.0002)	
if_bb_j	0.0001	
	(0.0002)	
log_bb_i		−0.0004***
		(0.0001)
log_bb_j		0.0002
		(0.0001)
log_NumPatents_i	0.010***	0.010***
	(0.0001)	(0.0001)
log_NumPatents_j	0.006***	0.006***
	(0.0001)	(0.0001)
log_Distance	−0.017***	−0.017***
	(0.0002)	(0.0002)
log(pop_i)	−0.001***	−0.001***
	(0.0001)	(0.0001)
log(pop_j)	−0.0002*	−0.0002
	(0.0001)	(0.0001)
log(income_i)	−0.00002	0.0004
	(0.0005)	(0.0005)
log(income_j)	0.007***	0.007***
	(0.0005)	(0.0005)
proximity	−0.021***	−0.021***
	(0.0003)	(0.0003)
Observations	977,751	977,751
R^2	0.042	0.042
Adjusted R^2	0.042	0.041
Residual Std. Error (df = 977732)	0.097	0.097
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 5: Dependent Variable: Citation Volume (IV Regressions)

	Dependent variable:			
	log_NumCitations			
	IV1: Adjacent Avg Broadband		IV2: Terrain Ruggedness	
	(1)	(2)	(3)	(4)
‘if_bb_i(fit)‘	0.004***		0.029***	
	(0.0004)		(0.009)	
‘if_bb_j(fit)‘	0.004***		0.028***	
	(0.0005)		(0.007)	
‘log_bb_i(fit)‘		−0.001***		−0.007***
		(0.0002)		(0.002)
‘log_bb_j(fit)‘		−0.0001		0.003
		(0.0002)		(0.003)
log_NumPatents_i	0.011***	0.011***	0.011***	0.011***
	(0.0002)	(0.0002)	(0.0002)	(0.0003)
log_NumPatents_j	0.006***	0.006***	0.006***	0.006***
	(0.0002)	(0.0002)	(0.0002)	(0.0001)
log_Distance	−0.017***	−0.017***	−0.018***	−0.017***
	(0.001)	(0.001)	(0.001)	(0.001)
log(pop_i)	−0.002***	−0.001***	−0.006***	−0.003***
	(0.0002)	(0.0002)	(0.001)	(0.001)
log(pop_j)	−0.001***	−0.0002	−0.005***	0.001
	(0.0001)	(0.0002)	(0.001)	(0.001)
log(income_i)	−0.003***	−0.001**	−0.012***	−0.002**
	(0.001)	(0.001)	(0.003)	(0.001)
log(income_j)	0.003***	0.005***	−0.005*	0.007***
	(0.001)	(0.001)	(0.003)	(0.002)
proximity	−0.021***	−0.021***	−0.021***	−0.021***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.106***	0.059***	0.369***	0.051***
	(0.020)	(0.017)	(0.077)	(0.016)
Observations	977,751	977,751	963,572	963,572
R^2	0.041	0.041	0.014	0.039
Adjusted R^2	0.041	0.041	0.014	0.039
Residual Std. Error	0.097 (df = 977740)	0.097 (df = 977740)	0.100 (df = 963561)	0.098 (df = 963561)
Note:				*p<0.1; **p<0.05; ***p<0.01

Changes in Voter Behavior Amidst Electoral Insecurity and Perceived Voter Fraud

April 2023

Abstract

The expansion of voting access through no-excuse absentee voting, early in-person voting, or mandatory vote-by-mail (VBM) during the COVID-19 pandemic has grown increasingly polarized along party lines in the United States. Widespread claims of fraudulent, missing, or otherwise miscounted ballots may have further eroded voter confidence in the nontraditional forms of voting mentioned above. Given the extensive nature of misinformation targeting mail-in voting and absentee voting, it is unclear how voters, particularly those affiliated with the Republican party, have responded and/or changed their preferred voting method in practice. This paper evaluates the impact of widespread allegations of fraudulent voting methods on individual-level voter behavior preceding and following the 2020 U.S. election cycle. Recent data collected from the North Carolina State Board of Elections details comprehensive voter registration data and voting history of over 4 million North Carolina residents spanning primary and general U.S. elections from 2016 to 2022. Utilizing a series of cross-sectional dynamic difference-in- differences analyses, we find that Republican voters were significantly less likely to vote by mail relative to Democratic or Independent voters following widespread allegations of fraudulent voting in early 2020 by prominent GOP leaders. However, the same subset of Republican voters was more likely to vote early absentee. These effects are more persistent among nonwhite Republican voters and in Republican-majority counties. As a robustness check for pre-treatment trends, an additional D.I.D. model with a fake placebo treatment further bolsters results. Taken altogether, these findings suggest that post-pandemic voter restrictions laws targeting nontraditional methods of voting may lead to changes in voter turnout and/or vote share along party lines.

Content

1 Introduction

Democratic backsliding of the United States, driven by political sectarianism, party polarization, and electoral insecurity, has accelerated in the past decade of American politics. Lockdowns during the COVID-19 pandemic motivated legislation that expanded voting access through no-excuse absentee voting, early in-person voting, or mandatory vote-by-mail (VBM), though these policies in question have grown increasingly polarized along party lines and subsequent widespread electoral misinformation have further eroded voter confidence. Former President Trump began casting doubt on the validity and trustworthiness of mail-in voting as early as April 2020, citing concerns of illegally printed ballots, forged voter signatures, and stolen ballots (Parks 2020; Saul and Epstein 2020). Despite these allegations, cases of voter fraud during U.S. elections have been exceptionally rare. Only 31 incidents of voter fraud between all U.S. elections from 2004 to 2014 have been conclusively discovered (Levitt 2014), during which over one billion votes were cast. With respect to mail-in voting, investigation of mail-in ballots in the 2016 and 2018 general elections found just 0.0025% rate of possible voter fraud (Viebeck 2020). Yet, the Republican Party has continued to oppose mail-in voting, early voting, or tighten restrictions on voting altogether—at least 17 states passed new restrictive voting laws in 2021 alone (Berry et al. 2021)—even though higher mail-in voting participation does not actually increase total Democratic voters relative to Republican voters, and hence, does not increase winning chances of a Democratic candidate (Elul et al. 2017; Atsusaka et al 2019; Thompson et al. 2020; Barber & Holbein 2020; Griffin 2021). Given these unsubstantiated claims against mail-in voting, it is unclear how voters, particularly those affiliated with the Republican party, have responded and/or changed their preferred voting method in practice. Could certain voters have rejected mail-in or early voting, options originally presented to simplify the voting process, in favor of the traditional, more arduous form of in-person voting? If so, which voters and for how long will this phenomenon persist?

Current literature on nontraditional voting methods is inadequate to answer these and related questions. Alleged fraud via mail-in voting is a relatively novel issue that gained extensive media coverage in 2020, and its real impacts on voting behavior have not been comprehensively studied. Despite ample evidence that refute claims of increased fraud via mail-in ballots, polling still show a stark partisan divide on the issue. Beyond ideological differences, few studies have analyzed actual election day

voting methods by party affiliation and voting privacy laws vary widely on a state-by-state basis and can further complicate analysis. Furthermore, questions regarding changes in voter behavior on an individual level over time in response to alleged voter fraud remain wholly unaddressed, and the few studies that analyze individual level voter behavior utilize only survey data rather than election data.

This paper intends to evaluate the impact of widespread allegations of fraudulent mail-in voting on individual-level voter behavior preceding and following the 2020 U.S. election cycle. Recent data collected from the North Carolina State Board of Elections details comprehensive voter registration and voting history of over 4 million North Carolina residents spanning primary and general U.S. elections from 2016 to 2022. Since VBM and absentee voting has been the target of alleged fraud by GOP leaders, one can reasonably hypothesize that Republican voters are rejecting nontraditional methods of voting in favor of in-person voting. Using a series of cross-sectional dynamic difference-in- differences analyses, we find a significant 6 percentage point decline in vote-by-mail rates among Republican voters relative to Democratic or Independent voters immediately following the 2020 primary and general elections. With regards to early absentee voting, Republican voters experienced a 5.2 percentage point increase in early voting rates, voting early at higher rates than Democratic/Independent voters. These effects are higher in magnitude for white Republican voters and Republican voters in Democratic-majority counties, with VBM rates on the 2020 general election experiencing a 13 percentage point and 9 percentage point decline, respectively. The dynamic difference-in-differences model relies on no pre-treatment trends between the control group (Democratic and Independent voters) and the treatment group (Republicans), known informally as the parallel trends assumption (Bertrand et al. 2004; Abadie 2005). No literature has conclusively correlated political party affiliation with a certain method of voting, nor does one expect the existence of concomitant shocks to disproportionately affect one voting group over the other. While the parallel trends assumption can thus be plausibly satisfied, robustness checks via additional estimators in Section 5 further corroborate this paper's findings.

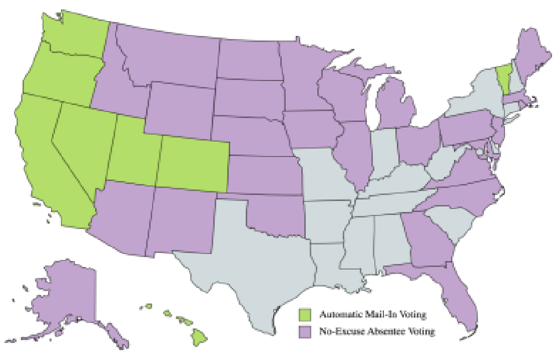
Following the Introduction, Chapter 2 discusses existing literature and empirical theory on voting methods and the effect of an increase in perception of voter fraud on voting behavior. Chapter 3 describes the dataset analyzed in this paper and methods of data cleaning and harmonization. Chapter 4 details the regression methodologies and approaches applied in this paper, followed by results and analysis of these findings in Chapter 5. Chapter 6 elaborates on the implications of these findings and suggests further avenues of research in this field.

2 Literature and Theory

2.1 Nontraditional Voting Methods

Traditional voting methods refer to voters who cast their ballots in-person on Election Day, whereas nontraditional voting methods refer to ballots submitted before Election Day and/or by mail via absentee ballots. Voting via nontraditional methods have surged in recent years, accounting for 69.4% of all voters in the 2020 presidential election, compared to 40.1% of voters in 2016. Upon closer examination, 43% of voters in the 2020 presidential election casted ballots by mail (VBM) and 26% voted in-person before Election Day (Scherer 2021). Absentee voting, sometimes referred to synonymously with nontraditional voting, refers to voting either by mail or before Election Day using an absentee ballot. Voters unable to cast their ballots in-person on Election Day may request an absentee ballot, though an increasing number of states have adopted "no-excuse" absentee voting, in which any voter can request and cast an absentee ballot without providing a reason or "excuse". As displayed in Figure 1, as of January 2023, 27 states and Washington D.C. offer "no-excuse" absentee voting, of which 8 states (California, Colorado, Hawaii, Nevada, Oregon, Utah, Vermont, and Washington) conduct elections entirely by mail—voters automatically receive a ballot in the mail without having to request one (National Conference of State Legislatures [NCSL] 2022).

Figure 1: Voting Methods By State



In recent years, voters who choose to use nontraditional methods of voting vary widely by age, education, sex, and race/ethnicity. In the 2020 presidential election, Scherer (2021) found that college-educated voters were more likely to vote by mail and/or before Election Day than voters with only a high school education. Similarly, 71.2% of women voted nontraditionally, compared to 67.4% of men. Voters identifying as Hispanic (76.7%) or Asian (82.4%) were significantly more likely to vote nontraditionally than non-Hispanic, white voters (67.5%). Unsurprisingly, voters aged 65 or over were more likely to utilize nontraditional methods of voting than their younger counterparts due to convenience.

At first glance, voters that elect to vote nontraditionally appear to align closely with those that tend to identify as Democrats. According to a 2021 Pew Research Center report, 17% of registered Republican voters identify as non-white as compared to 40% of Democratic registered voters. Further, the Republican Party has gained a foothold among white non-college voters, while college-educated voters are more likely to lean Democratic. Gender gaps in party identification are starker than ever before: 56% of women identify as Democratic, as compared to only 42% of men (Atske 2020). Atske (2020) also found the growing generational gap between Republican and Democrats is growing—54% of millennials identify as Democratic, as composed to 38% of millennials as Republicans.

In considering the crossovers between voter subgroups that are more likely to identify with the Democratic Party and those which are likely to vote nontraditionally, the recent surges in Republican-backed legislations to restrict VBM or early voting can be rationalized as attempts to restrict voting access to Democratic voters. However, as popularity of the Republican Party continues to decline among millennials, the convenience offered by nontraditional methods of voting to a rapidly aging Republican-dominated electorate may outweigh the risk of perceived fraud associated with these voting methods. As such, attempts to restrict voting access may backfire onto the Republican Party, as discussed in Section 2.2 – 2.3.

2.2 Vote-by-Mail Effects

The Democratic Party and Republican Party both prioritize free and fair elections, though execution towards this goal differs widely along partisan lines. Concerns regarding election integrity and voter fraud motivate Republican-backed legislation that place additional requirements on voting eligibility, while Democrats fear that these hurdles may discourage legitimate voters, or worse, disenfranchise eligible voters. As of May 2021, Republicans have introduced, filed, or passed over 250 bills across 48 states to restrict voting access (Berry et al. 2021). Of these, 125 bills target vote-by-mail or include provisions that further restrict VBM. Despite these efforts, nontraditional voting methods have been shown to not have a substantial impact on either political party. Analysis of the effects of universal VBM across elections from 1968 to 2018 in California, Utah, and Washington indicated no significant increase in either party's turnout or vote share (Hanmer and Traugott 2004; Elul et al. 2017; Thompson et al. 2020; Barber and Holbein 2020). While mandatory VBM has been shown to increase voter turnout among minority groups and younger voters, its effect on election outcome is unclear and has not been shown to increase the likelihood of a Democratic candidate winning (Atsusaka et al. 2019; Griffin 2021). Even in states without universal VBM, VBM does not increase voter turnout, nor does it make electorates more representative of the voting-age population (Berinsky et al. 2001; Yoder et al. 2021). Studies of the effects of a Texas state voting law that allows voters ages 65 or older to vote no-excuse absentee found that eligible no-excuse absentee voters were not more likely to vote absentee than their younger counterparts (Meredith and Endter 2016; Yoder et al. 2021).

Conversely, vote-by-mail may in fact harm the Democratic Party, particularly for voters who are frequent movers. For minorities, low-income groups, and younger voters who relocate at a greater frequency than their white, high-income, older counterparts, ballots received by mail may be missent to former addresses, particularly for voters who may not update their registration often (Wines 2020). Additionally, mail-in ballots are rejected at a higher rate than in-person ballots, often due to erroneous or illegible marks. Among younger voters and Hispanic voters, rejection rates in the 2018 Florida general election were substantially higher (Baringer et al. 2020). Overall, mail-in ballots by newly registered voters, inexperienced voters, minorities, and female voters are more likely to be rejected (Cottrell et al. 2021; Shino et al. 2022).

2.3 Early Voting Effects

Nontraditional methods of voting also encompass in-person voting completed before Election Day. As of August 2022, only 4 U.S. states—Alabama, Connecticut, Missouri, and New Hampshire—do not offer early voting, while the remaining 46 states, including the District of Columbia,

American Samoa, Guam, Puerto Rico, and the Virgin Islands, all offer some form of in-person voting before Election Day (NCSL 2022).

Ranging from local/state officials to former President Donald Trump, Republican opposition to expansion of voting access has extended to early voting as well. Famously, Texas Republican Party Chairperson Allen West, alongside other key Republican officials, sued Texas Governor Greg Abbott in 2020 over his expansion of early voting amidst the coronavirus pandemic—the case was later rejected by the state Supreme Court (Mulder 2020). Effects of permitting early voting on voter turnout are generally positive (Herron and Smith 2014; Glynn and Kashin 2017; Kaplan and Yuan 2020). Several studies point to a more drastic increase in turnout for women and minority voters (Herron and Smith 2014; Kaplan and Yuan 2020). However, early voting expansion in the early 2000's has been shown to have either a negligible impact on voter turnout, or in some cases, paradoxically decreased turnout (Gronke et al. 2007; Burden et al. 2014). Burden et al. (2014) theorizes that early voting stretches out the period in which voters can drop their ballots, thereby reducing the likelihood of being seen voting by other eligible voters, which can be an important determinant of voting as documented by Dellavigna et al. (2016). It should be noted that a positive relationship between voter turnout and Democratic success has not been conclusively identified, although studies of social democratic vote share in other countries with two major political parties have indicated the possibility of this correlation (McAllister 1986; Nagel 1988). These empirical findings conflict with more recent findings from Fisher (2007), concluding that higher turnout does not causally affect the left share of votes across 23 OECD countries from 1960 to 2002.

2.4 Perceived Voter Fraud

Despite widespread allegations of increased voter fraud due to VBM in the 2020 U.S. primary and general elections, cases of voter fraud remain exceptionally rare. Analysis of over 1 billion ballots dropped across all general, primary, special, and municipal elections from 2000 to 2014 found only 31 incidents of fraud (Levitt 2014). As an increasing number of states implement universal mail voting, no evidence has linked risk of voter fraud to VBM (Viebeck 2020; Auerbach and Pierson 2021). In some cases, fraud rates have been estimated to decrease due to mandatory VBM. Auerbach & Pierson (2021) estimates that 73 additional cases of voter fraud across Washington elections held from 2011 to 2019 would

would have been reported had the state not implemented mandatory VBM.

Debunking claims of increased fraud due to VBM has done little to diminish the stark partisan divide on the issue. The majority of voters who supported Trump in the 2020 U.S. Presidential Election believe election fraud was widespread, particularly among voters who, paradoxically, were more politically aware and closely following election news (Lockhart et al. 2020; Pennycook and Rand 2021). This phenomenon is best explained by Benkler et al (2020), concluding that concerns over voter fraud were primarily driven by the political elites and mass media rather than social media. The corrosive effects of widespread and unsubstantiated claims of voter fraud go beyond the delegitimization of election results and have been shown to reduce voter confidence in electoral integrity (Berlinski et al. 2021). As of April 2021, Republican voters are four times more likely to support restricting early or mail-in voting only to voters with a valid documented reason as compared to Democratic voters (Nadeem 2022). Disturbingly, corrective measures by media sources such as fact-checking have not been shown to mitigate the increasingly extremist views against the prevalence of perceived fraud. (Benkler et al. 2020; Berlinski et al. 2021).

Beyond ideological differences, few studies have analyzed actual election day voting methods by party affiliation, instead relying on survey data rather than exit polling or state election data. The following section, Chapter 3, describes and presents summary statistics on the statewide, individual-level voter data retrieved from the North Carolina State Board of Elections analyzed in this paper.

3 Data

Few studies, if any, use individual-level voter registration and voter history data to compare voter behavior preceding and following the 2020 U.S. primary and general elections, and those which do so opt instead to use data collected from exit polling, survey experiments, and/or media stories. States vary widely in the degree to which voter registration data is publicly available, and if available, often come with high costs, restrictions, and/or limitations. North Carolina remains as one of the few states that regularly maintain voter registration records and historical voting history data of its residents. The data sample used in this paper is retrieved from the North Carolina State Board of Elections, consisting of

all voting history and registration data of all 100 counties in North Carolina. The following sections detail key categorical and control variables included in the dataset, followed by data harmonization and cleaning in preparation for analysis.

3.1 Voter Registration Records

The North Carolina State Board of Elections (NCSBE) maintains a comprehensive and public database of voter registration records, updated weekly, across over 15 years. Current voter-level registration records and snapshots of voter registration records contain all individuals currently registered or formerly registered to vote in North Carolina by county. Records include registration status, voter demographics (race, ethnicity, age, and gender), party affiliation, mailing address, and crucially, each voter's unique North Carolina Identification Number (NCID). These voter registration numbers can be linked to voter history data files, discussed in Section 3.2. Party affiliation is recorded under either "Republican", "Democratic", or "Unaffiliated" categories. Race is recorded as indicator variables under white, Black, Asian Americans and Pacific Islander, Native Indian, Mixed, and Other.

NSCBE does not record individual household income, birthdays, Social Security number, or driver's license number, as they are confidential under state law. Apart from income, these remaining omitted figures are not necessary for the purposes of this paper. As income may be relevant and potentially confounding, city-level fixed effects can partially control for differences in voter income. All other relevant demographic data are represented as indicator variables.

3.2 Voter History

Voter-level history data, also maintained by the North Carolina State Board of Elections (NCSBE), contains voter-level, weekly-updated data for every North Carolinian election over the past 10 years, as well as group-level, voter demographic counts for each election over the last past two decades. These datasets contain a data entry for each election in which a voter has participated in the past 10 years, including NCID, party affiliation, county, precinct, and most importantly, voting method.

Voting method is recorded under seven possible categories: *In-person, Absentee by Mail, Absentee One-Stop, Curbside, Absentee Curbside, Pro-*

Provisional, and Transfer. In-person refers to ballots cast on Election Day, in-person, at accepted voting locations. Absentee by Mail refers to ballots returned by mail, while Absentee One-Stop refers to ballots returned in-person before Election Day at either the county board of elections office or at one-stop early voting sites. Curbside and Absentee Curbside voting options are reserved for voters, who due to age or disability, are unable to enter voting sites to vote in-person or early, respectively. *Provisional* ballots are provided in-person to voters when additional detail is needed to ensure eligibility to vote. *Transfer* voting refers to transfer ballots dropped in-person at various precincts on Election Day by voters who failed to update their address by the voter registration deadline.

3.3 Data Cleaning and Harmonization

Data from all primary and general elections, totaling eight elections, held on even- numbered years from 2016 to 2022, were examined for the purposes of this paper; voting data on all other local, municipal elections were filtered out to ensure standardization and applicability of analysis. Elections held in North Carolina prior to 2016 were subject to a controversial voter ID law passed in 2013, requiring all voters to show photo ID, cut early voting by a week, mandated all voters to vote in their assigned precinct, and prohibited voters from registering and voting on the same day. The 4th Circuit Court of Appeals overturned this law in 2016, arguing that the law was designed with "discriminatory intent" and disproportionately barred African-Americans from voting (Domonoske 2016).

As this paper intends to measure changes to voter methods prior to widespread misinformation about mail-in ballot fraud in 2020, referred to hereafter as the "treatment", voter registration data were filtered to only include voters who had voted in at least one election "pre-treatment" and one election "post-treatment". Doing so ensures that the remaining pool of voters were eligible to participate in both pre- and post-treatment elections. Additional filters applied to the remaining pool of voters eli inated voters who only turned of age following the 2020 elections as well as voters who have passed away or relocated since 2020. The primary election, held on March 3, 2020, is categorized as the first "post-treatment" election, as allegations of voter fraud via mail-in ballots began to ramp up in the early spring of 2020 into the presidential election in the fall. To avoid multicollinearity, the time variable denoting the last "pre-treatment" election is omitted and normalized to 0, and subsequent analysis results

are interpreted relative to this election (see Section 4 for details).

Construction of the final dataset required merging voter history and voter registration records using individual voter registration number from the remaining pool of voters that meet the criteria above. Additional indicator variables were generated for race, gender, political party affiliation, and elections type (primary election or general election). Using a series of dynamic difference-in-differences regressions, another binary variable indicated if elections were held prior or following 2020, with the primary election of 2020 set as the first post-treatment election.[1]

Efforts to restrict mail-in ballots have already culminated in legislation in several states. North Carolina Senate Bill 326, introduced in 2020, known as the Election Day Integrity Act, would have reduced the time to request and return absentee ballots, and would have invalidated over 30,000 absentee ballots from 2020 elections (Election Day Integrity Act 2021). However, while the bill passed the state senate's Redistricting and Elections Committee in June 2021, North Carolina Governor Cooper ultimately vetoed the bill in December of that year — the bill never went into effect.

As such, the scope, size, and methodology of this study should widen the generalizability of voter behavior to most states in the United States, by controlling for demographic data and differences between presidential year elections and non-presidential year elections. Additionally, it should be noted that elections in 2020 may be affected by external factors due to COVID-19, which may be controlled for via difference-in-differences regression method that compares voting behavior along party lines.

3.4 Summary Statistics and Raw Trends

Table 1 displays the descriptive statistics of the complete, merged dataset by political party affiliation across all elections from 2016 to 2022, subject to the criteria detailed in the previous sections. The dataset accounts for over 4.4 million North Carolinian voters who have voted in at least one pre-treatment election (before 2020) and one post- treatment election (2020 onwards). While North Carolina allows voters to select"Unaffiliated" under party preference during voter registration, the number of voters identifying as Democratic only slightly outnumber Republican voters and are otherwise relatively equivalent. Key differences in voter

1 See Section 3.4 for description and summary statistics of relevant variables.

behavior along party lines can be seen in the voting methods category, where Republicans are more likely to vote in-person than Democratic voters (96.1%) and are subsequently less likely to vote early (50.4%) or by mail (3.9%)[2].

As expected, fundamental differences in party affiliation of demographic groups are displayed along racial, gender, and age groups. Notably, 94.2% of Republican voters identified as white, as compared to only 45.2% of Democratic voters. In contrast, only 10% of Republican voters were African-American, as compared to 46.3% of Democratic voters. Gender gaps across party lines were also highlighted—women were more likely to vote consistently in both pre- and post-treatment elections than men, making up 53.5% of voters in the dataset as compared to the 42.9% of voters who identified as male. Gender differences in partisan identification were equally prevalent: 60.6% of Democratic voters were women, as compared of men's share of 36.3%.

Table 1: Summary Statistics By Party Affiliation

Variables	All voters (n = 4,402,220)		Republican (n = 1,455,725)		Democratic (n = 1,588,387)		Unaffiliated (n = 1,340,480)	
Voting method								
In-person	0.944	(0.229)	**0.961**	(0.193)	**0.935**	(0.247)	0.937	(0.243)
Early voting	0.525	(0.499)	**0.504**	(0.500)	**0.555**	(0.497)	0.512	(0.500)
By mail	0.056	(0.229)	**0.039**	(0.193)	**0.065**	(0.246)	0.063	(0.243)
Absentee voting*	0.581	(0.493)	0.542	(0.498)	0.621	(0.485)	0.575	(0.494)
Age	56.97	(17.02)	58.97	(16.38)	57.95	(17.28)	53.79	(16.90)
Race								
White	0.717	(0.450)	0.942	(0.233)	0.452	(0.498)	0.786	(0.410)
Black	0.199	(0.400)	0.100	(0.100)	0.463	(0.499)	0.094	(0.292)
AAPI	0.011	(0.106)	0.006	(0.074)	0.100	(0.100)	0.019	(0.137)
Native	0.006	(0.079)	0.004	(0.063)	0.009	(0.093)	0.006	(0.077)
Other	0.061	(0.240)	0.037	(0.188)	0.060	(0.237)	0.089	(0.285)
Two or more	0.005	(0.067)	0.002	(0.041)	0.006	(0.076)	0.006	(0.077)
Hispanic	0.020	(0.140)	0.010	(0.097)	0.025	(0.155)	0.025	(0.157)
Gender								
Male	0.429	(0.495)	0.475	(0.499)	0.363	(0.481)	0.455	(0.498)
Female	0.535	(0.499)	0.499	(0.500)	0.606	(0.489)	0.492	(0.500)
Undisclosed	0.036	(0.186)	0.026	(0.158)	0.031	(0.173)	0.053	(0.224)

**Note: Absentee voting refers to any votes submitted using an absentee ballot, including mail-in voting and early voting. Data reported spans primary and general elections from 2016 – 2022.*

2 Recall that early voting in North Carolina is known as "Absentee Onestop" or "Absentee Curbside". "Absentee by Mail" refers to ballots dropped by mail. All other voting options are considered voting in-person. See Section 3.2 for details.

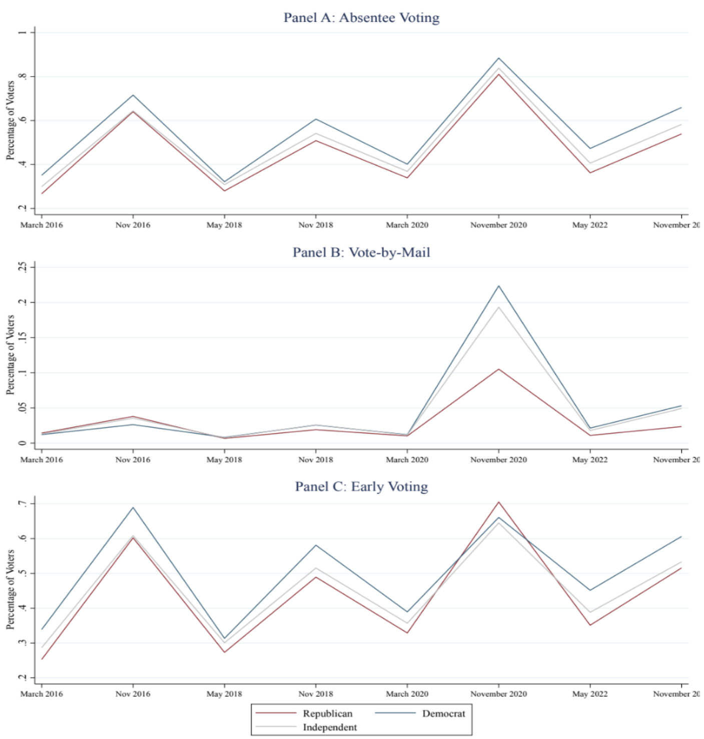
Table 2 displays descriptive statistics of voting methods for each election year from 2016 to 2022. Each column reports the mean of each variable across both the primary and general elections of each year. Except for the 2020 primary and general elections, VBM only makes up between 2 to 4% of submitted ballots, far outnumbered by ballots submitted in-person.. The jump of VBM in 2020 can be attributed to the start of COVID-19 pandemic, as further highlighted by Figure 2. The impact of the pandemic complicates regression analysis particularly in the case of difference-in-differences methodology, as the control group's (Democratic and/or Independent voters) voting behavior should ideally remain consistent over time. Despite this, the difference-in-differences analysis focuses on changes in voting behavior across political parties that were all affected by COVID-19. Thus, Democrats and Independents serve as an effective and useful control group when analyzing the effect of widespread electoral insecurity following the 2020 elections. Interestingly, VBM lowered to near pre-pandemic levels immediately in the primary election of 2022. Additionally, VBM seems to be a more popular voting method among both Democratic and Republican voters for general elections as compared to primary elections, which will be accounted for via indicator variables for election type.

Table 2: Summary Statistics By Election Year

Variables	2016 Elections (n = 6,059,585)		2018 Elections (n = 4,416,024)		2020 Elections (n = 6,323,139)		2022 Elections (n = 4,359,741)	
Voting method								
In-person	0.974	(0.160)	0.980	(0.141)	0.877	(0.328)	0.966	(0.182)
Early voting	0.520	(0.499)	0.483	(0.500)	0.573	(0.497)	0.506	(0.500)
By mail	0.026	(0.160)	0.020	(0.141)	0.123	(0.328)	0.034	(0.182)
Absentee voting*	0.546	(0.498)	0.503	(0.500)	0.696	(0.460)	0.540	(0.498)
Age	59.11	(16.52)	59.66	(16.64)	58.05	(16.86)	59.91	(15.67)
Race								
White	0.734	(0.442)	0.729	(0.444)	0.718	(0.450)	0.756	(0.429)
Black	0.196	(0.397)	0.201	(0.400)	0.204	(0.403)	0.181	(0.385)
AAPI	0.009	(0.092)	0.009	(0.096)	0.102	(0.100)	0.008	(0.090)
Native	0.006	(0.077)	0.007	(0.082)	0.006	(0.078)	0.005	(0.073)
Other	0.052	(0.240)	0.051	(0.219)	0.057	(0.232)	0.046	(0.209)
Two or more	0.003	(0.062)	0.004	(0.063)	0.004	(0.066)	0.003	(0.058)
Hispanic	0.016	(0.125)	0.016	(0.124)	0.018	(0.133)	0.012	(0.109)
Gender								
Male	0.429	(0.495)	0.437	(0.496)	0.429	(0.495)	0.442	(0.497)
Female	0.541	(0.498)	0.534	(0.499)	0.538	(0.499)	0.530	(0.499)
Undisclosed	0.030	(0.171)	0.029	(0.169)	0.034	(0.180)	0.028	(0.166)

**Note: Absentee voting refers to any votes submitted using an absentee ballot, including mail-in voting and early voting.*

Figure 2: Nontraditional Voting Methods Over Time, 2016-2022



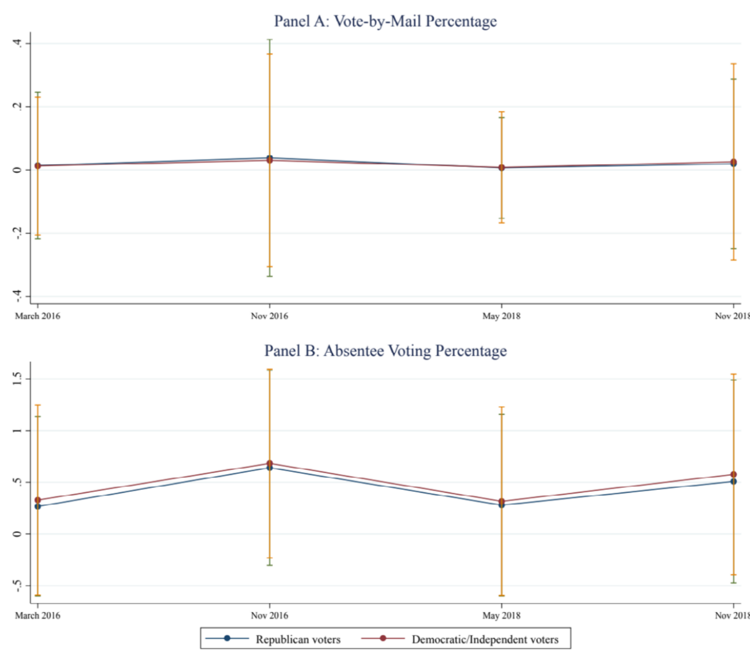
4 Empirical Strategy

Depending on the variable of interest, several models can be run using the cleaned and filtered voter registration and history data. Given that this paper is primarily interested in examining changes in voting method pre- and post-treatment based on party affiliation, a cross-sectional difference-in-differences event study logistic regression is best suited for this analysis. Consistent with other studies examining changes in voting method over time, a difference-in-differences method best compares voting behavior preferences over time among varying voter demographic groups (Barber and Holbein 2020; Yoder et al. 2021). While Yoder et al. (2021) and Barber and Holbein (2020) use the difference-in-differences design to estimate voter turnout pre- and post-treatment years, this paper capitalizes on individual-level voting method data to estimate changes in nontraditional voting along party lines over time. By normalizing

the election immediately preceding treatment year (2018 general elections) to zero, resulting analysis displays the difference in effects of treatment on Republican voting behavior as compared to Democratic voters relative to the spring of 2020.

Difference-in-differences estimation requires two major assumptions: the parallel trends assumption and Stable Unit Treatment Value Assumption. The former necessitates parallel trends in the outcomes of treatment groups and control groups in the absence of treatment (Bertrand et al. 2004; Abadie 2005). In the context of this paper, the difference in voting method preference among Republican and Democratic voters should remain constant over time, absent of the widespread allegations of fraud by key Republican leaders in 2020. Figure 3 displays the average proportion of voters utilizing VBM and absentee voting (VBM and early voting) in pre-treatment elections by political party affiliation. Parallel trends can be seen in absentee voting (Panel B), though the parallel trends assumption is violated when examining VBM alone. This is likely due to the smaller sample size of voters who elect to vote-by-mail—about 5% on average across all elections, as seen in the orange and green standard error bars.

Figure 3: Parallel Trends, Nontraditional Voting Methods



On this note, since estimation of the aforementioned models in this paper relies on individual level, state-wide data spanning 2016 to 2022, the sample sizes in these estimations frequently exceed a couple million entries. Hill and Leamer (1980) established that as sample size grows, classical hypothesis testing at a fixed level, usually at the 5% significance level by convention, may distort results and increase likelihood of Type I errors-in which a true null hypothesis is falsely rejected. Hill and Leamer thus establish an inverse relationship between sample size and significance level, where a more accurate significance is the square root of the natural logarithm of sample size ($\sqrt{\ln(N)}$). As an example, for a sample size of 14 million, a null hypothesis with a t-statistic exceeding 4.06 should reasonably be rejected, as opposed to the conventionally used value, 1.96, at the95% confidence interval. While most widely-used econometric software still relies on fixed level hypothesis testing by convention, statistical significance as reported in the regression output in the subsequent results in Chapter 5 may not be the most reliable measure of significance or proof of pre-treatment trends. Rather, one should examine the magnitude and direction of the estimated interaction coefficients in conjunction with conventional fixed hypothesis testing (at 1%, 5%, and 10%) as a measure of statistical significance and determination of negligibility.

The Stable Unit Treatment Value Assumption (SUTVA) requires that outcomes of treatment groups depend only on assigned treatment, rather than spillover effects from external factors/treatment. While no satisfactory statistical solutions exist to test SUTVA, spillover effects can best be minimized by reasonably defining experimental units and control variables. A corollary of these assumptions requires control and treatment groups display no anticipation of treatment. On this note, anticipation of voter fraud is unlikely for both Republican and Democratic voters, as claims of voter fraud were accelerated primarily by the primary and general election of 2020 by former President Trump and could not have been foreseen by voters.

4.1 Pooled Difference-in-Differences Estimation

The main outcome of interest in analyzing voter behavior is whether Republican voters are employing one certain voting method at varying rates as compared to Democratic voters. The difference-in-differences ordinary least squares estimator compares the disparity in voting method preference by party affiliation pre- and post-primary election of 2020. The following specification details the model employed and relevant variables of interest:

$$(1) \qquad y_{i,t} = \beta_0 + \beta_1 \cdot PostTreatment_{i,t} + \beta_2 \cdot (Party \cdot PostTreatment)_{i,t} + \beta_m \cdot Z_{i,t} + \delta_m + \varepsilon_{i,t}$$

Equation (1) models the pooled difference-in-differences ordinary least squares estimation, where $y_{i,t}$ is an indicator variable for the tested voting method (by mail, early voting, or both) for individual *i* in election *t*; $PostTreatment_{i,n}$ is an indicator variable for elections on or after 2020; $Party_i$ is an indicator variable for the political party of which individual voter is registered with; $Z_{i,t}$ accounts for individual characteristics including gender, age, race, and Hispanic origin; δ_m accounts for location-based fixed effects, controlling for time-invariant characteristics specific to certain geographic areas by zip code, such as income. The main variables of interest are β_1 and β_2. β_1 signals changes in voting method preference in elections on or after 2020 across voters in the control group (whichever political party affiliation that is not represented by $Party_i$). The inclusion of individual and election-year fixed effects controls for static differences in outcome between voters as well as any time-varying, regional shocks.

The main outcome of interest is β_2, which measures change in voting method preference in the treatment group beginning in 2020. For example, a positive β_2 signals an increase in voting preference for the tested voting method ($y_{i,n}$), and vice versa.

4.2 Paneled Difference-in-Differences Estimation

The pooled difference-in-differences approach discussed above treats pre-treatment elections and post-treatment elections as binary without accounting for year-to-year fixed effects or changes in voter behavior across individual elections. While offering a simpler interpretation, the alternative, a paneled cross-sectional difference-in-differences estimator can capture year-to-year changes in voting method across party lines:

$$(2) \qquad y_{i,t} = \beta_0 + \sum_{j=1}^{8} \beta_j \cdot Party_i \cdot Election_{j=t} + \sum_{j=9}^{16} \beta_j \cdot Election_{j=t+8} + \beta_k \cdot Party_i + \beta_m \cdot Z_{i,t} + \delta_m + \varepsilon_{i,t}, \qquad \text{where } k \neq m \neq j.$$

Equation (2) details the paneled cross-sectional difference-in-differences ordinary least squares estimation, where $y_{i,t}$ is again an indicator variable for the tested voting method (by mail, early voting, or both) for individual i in election t; β_j is a series of difference-in-differences coefficients estimating the effect of post-treatment elections alone and on Republican voters; $Election_j$ is an indicator for each election t; $Z_{i,t}$ again accounts for individual characteristics including gender, age, race, and Hispanic origin and δ_m likewise accounts for location-based fixed effects, controlling for time-invariant characteristics specific to certain geographic areas by zip code. To avoid collinearity, the first pre-treatment indicator ($j = 4$) is forced to zero. Assuming all primary and general elections from 2016 to 2022 are included in the analysis ($1 \leq t \leq 8$), the main parameters of interest are β_5 to β_8 (the difference-in-differences post-treatment estimators), and the remaining coefficients β_9 to β_{17} are the estimates of the linear terms.

Note that the model expressed in Equation (2) is flexible and can be adapted with or without varying levels of fixed effects and individual-level characteristics. The outcome variable of interest $y_{i,n}$ can represent likelihood of voting-by-mail (VBM), early voting, or both, a.k.a. nontraditional voting methods. Similarly, the indicator variable for political $Party_i$ can return a true value of 1 for either Republican voters, but also Democratic voters, or unaffiliated voters. Indicator variables for election type (primary vs. general elections, presidential vs. midterm years) can be added to account for election-type varying characteristics but were purposefully not accounted for in (1) to avoid unneeded complexity.

5 Results

5.1 Do Republican voters vote less nontraditionally?

Nontraditional voting, to review, broadly covers any voting method that differs from in-person voting on Election Day, including early voting, vote-by-mail, and early curbside voting. North Carolinian voters who vote nontraditionally only do so using absentee ballots, as North Carolina is a no-excuse absentee voting state; hereinafter, absentee voting is used synonymously with nontraditional voting. Table 3 displays the outcome coefficients from running the pooled difference-in-differences

regression on pre- and post-treatment elections.

Table 3: Pooled Difference-In-Differences Absentee Voting

Table 3: POOLED DIFFERENCE-IN-DIFFERENCES, ABSENTEE VOTING

Variables	(1)	(2)	(3)	(4)
REP	-0.0793***	-0.0827***	-0.0786***	-0.0638***
	(0.00510)	(0.00592)	(0.00397)	(0.00703)
POST	0.100***	-	-	-
	(0.00277)			
POST × REP	0.000518	-0.00812***	-0.00704***	-0.00701***
	(0.00316)	(0.00230)	(0.00233)	(0.00227)
WHI				-0.0188***
				(0.00264)
AAPI				0.0152***
				(0.00516)
BLA				0.0175***
				(0.00533)
NAT				-0.0670***
				(0.00978)
MIX				0.0247***
				(0.00373)
HISP				-0.0330***
				(0.00293)
MALE				-0.0157***
				(0.000631)
LN_AGE				0.193***
				(0.00747)
Constant	0.569***	0.623***	0.621***	-0.152***
	(0.00690)	(0.00620)	(0.00162)	(0.0328)
Election FE	NO	YES	YES	YES
Zip Code FE	NO	NO	YES	YES
Demographic Controls	NO	NO	NO	YES
Observations	14,392,277	14,392,277	14,392,277	14,392,277
R-squared	0.017	0.142	0.162	0.176

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (1), with Republicans as treatment group and Democrats as control group. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age)

Following the widespread allegations of voter fraud via absentee ballots in 2020, the average Republican voter was only slightly less likely to vote nontraditionally as compared to Democratic voters, though this effect is negligible, at roughly a 0.7 percentage point decline. A point of interest is the estimated coefficients on *Post* and *Rep*. Regression (1) estimates Equation (1) without individual election fixed effects, where *Post* is an indicator for all elections on or after 2020. In this model, there is an e timated 10 percentage point increase in nontraditional voting in the post-treatment elections, possibly hinting at effects of the COVID-19 pandemic on voting behavior regardless of political party affiliation. This estimate is then omitted due to collinearity in regressions (2) – (4) following the incorporation of election fixed effects. Across all estimations, the coefficient on *Rep* hovers between a 6.8 to 8.3 percentage point decline in voting absentee throughout all election periods. These estimates reinforce the general hesitation among Republican voters to vote absentee, though these effects do not identify the specific voting methods of which Republican voters are underutilizing, though significant at the 99th confidence interval. Repeating this estimation on more specific voting methods would likely be more fruitful, as displayed in Sections 5.2 onwards.

An additional note on the estimates of demographic controls—non white and non- Hispanic voters are generally significantly more likely to vote nontraditionally, particularly among Asian-American Pacific Islander and African-American voters. White, male voters are less likely to vote nontraditionally. As expected, age is by far the greatest demographic predictor of absentee voting, with an estimated 19.4 percentage point increase in likelihood for each additional one percent increase in age. These effects corroborate findings by Scherer (2021).

Table 4 displays the estimation coefficients from running the paneled cross-sectional difference-in-differences regression on pre- and post-treatment elections as in Equation (2)[3].

Rather than pooling all pre-treatment and post-treatment elections together, an indicator variable is generated for each election, and the average treatment effect is measured by interacting these indicator variables with treatment variable Rep. Election year indicator variables beginning with “P” signals a primary election while “G” indicates a general election. The last two numeric characters signal the year in which the election held (e.g. P20 is the indicator variable for the primary election of 2020).

3 For full table of estimation coefficients, see Appendix Table A1.

Table 4: Dynamic Difference-In-Differences Absentee Voting

Variables	(1)	(2)	(3)
REP	-0.0764***	-0.0729***	-0.0566***
	(0.00695)	(0.00573)	(0.00903)
REP × P16	-0.00823	-0.00628	-0.00822
	(0.00576)	(0.00551)	(0.00576)
REP × G16	-0.0230***	-0.0223***	-0.0264***
	(0.00361)	(0.00355)	(0.00358)
REP × P18	0.0346**	0.0368***	0.0391***
	(0.0141)	(0.0130)	(0.0135)
REP × P20	0.0135**	0.0177***	0.0107**
	(0.00592)	(0.00527)	(0.00524)
REP × G20	0.00126	0.00138	-0.000393
	(0.00392)	(0.00392)	(0.00388)
REP × P22	**-0.0342*** **	**-0.0288*** **	**-0.0263*** **
	(0.00764)	(0.00716)	(0.00692)
REP × G22	**-0.0438*** **	**-0.0425*** **	**-0.0421*** **
	(0.00513)	(0.00504)	(0.00510)
P16	-0.256***	-0.257***	-0.267***
	(0.00614)	(0.00598)	(0.00606)
G16	0.110***	0.111***	0.111***
	(0.00436)	(0.00428)	(0.00427)
P18	-0.285***	-0.282***	-0.300***
	(0.0115)	(0.0121)	(0.0123)
P20	-0.206***	-0.208***	-0.213***
	(0.00504)	(0.00476)	(0.00479)
G20	0.279***	0.280***	0.284***
	(0.00691)	(0.00666)	(0.00658)
P22	-0.133***	-0.135***	-0.153***
	(0.00850)	(0.00828)	(0.00809)
G22	0.0542***	0.0540***	0.0498***
	(0.00330)	(0.00334)	(0.00341)
Constant	0.605***	0.602***	-0.167***
	(0.00729)	(0.00336)	(0.0333)
Zip Code FE	NO	YES	YES
Demographic Controls	NO	NO	YES
Observations	14,392,277	14,392,277	14,392,277
R-squared	0.143	0.163	0.177

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (2), with Republicans as treatment group and Democrats as control group. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age).

Following the widespread allegations of voter fraud by Republican party leaders in 2020, average rates of nontraditional voting dropped among Republican voters in the primary and general elections of 2022. These effects are modest and positive at first, with a roughly 1 percentage point difference in the primary election of 2020, peaking to over a 4.2 percentage point decline in the general election of 2022. It should be noted results from Table 4 should be taken with caution, since interaction coefficients for certain pre-treatment covariates are significant, potentially signaling pre-treatment trends, though more likely due to the overwhelmingly large sample size (N = 14,392,277). These assumptions will be further examined in Section 5.4 – 5.5 as further robustness checks.

As there have only been four post-treatment elections as of March 2023, the persistence of these effects is unclear and don't seem to level off at any election year. Estimates on the linear election year indicator variable for the general election of 2020 (*G20*) are drastically higher as compared to other election year indicator variables, likely attributed to the effects of the COVID-19 pandemic. Demographic coefficients are not displayed in Table 4 for brevity and to avoid redundancy, though the effects are similar in direction and magnitude as in Table 3.

While the estimators displayed in Table 4 provide more insight into the effects of widespread misinformation on voting methods, absentee voting is a broad term that accounts for both vote-by-mail (VBM) and early voting. To breakdown these effects, specific analysis on each voting method is warranted and displayed in the following sections.

5.2 Do Republicans voters vote less by mail?

Because VBM has become a target of unsubstantiated claims of fraud by the Republican Party, we expect Republicans to vote less by mail as compared to Democrats/Independents following the 2020 elections. Table 5 displays the outcome coefficients from running the pooled difference-in-differences regression on VBM pre- and post-treatment elections.

Regression analysis confirms that Republican voters are less inclined to vote-by- mail beginning in the general election of 2020 as compared to Democratic voters. These results indicate that Republican voters experienced an estimated 5.8 – 6.0 percentage point decline in VBM rates beginning in the 2020 elections. In considering the effects of the COVID-19 pandemic, which is assumed to be an exogenous shock affecting all voters, one would expect voters to utilize nontraditional methods of voting more frequently, as voting-by-mail or voting early is far more convenient and time-saving as compared to voting in-person on Election Day. As such, the higher propensity for voting-by-mail following the 2020 elections among Democratic voters is unsurprising, yet remain in stark contrast to Republican voters, who consistently show a decrease in likelihood of voting by mail or voting absentee. Interestingly, nonwhite, male

voters are less likely to vote by mail. As expected, the age of voters directly and positively correlates with the likelihood of voting-by-mail.

Table 5: Pooled Difference-In-Differences, Vote-By-Mail

Variables	(1)	(2)	(3)	(4)
REP	0.00279***	0.00252***	0.00201*	-0.0112***
	(0.000940)	(0.000950)	(0.00111)	(0.000884)
POST	0.0840***	-	-	-
	(0.00259)			
POST × REP	-0.0576***	-0.0599***	-0.0591***	-0.0589***
	(0.00190)	(0.00216)	(0.00216)	(0.00217)
WHI				-0.0162***
				(0.00227)
AAPI				0.0238***
				(0.00278)
BLA				-0.0519***
				(0.00209)
NAT				-0.0314***
				(0.00229)
MIX				-0.00809***
				(0.00180)
HISP				-0.0169***
				(0.00179)
MALE				-0.00517***
				(0.000250)
LN_AGE				0.0398***
				(0.00197)
Constant	0.0205***	0.0639***	0.0639***	-0.0647***
	(0.000949)	(0.00234)	(0.000699)	(0.00871)
Election FE	NO	YES	YES	YES
Zip Code FE	NO	NO	YES	YES
Demographic Controls	NO	NO	NO	YES
Observations	14,392,277	14,392,277	14,392,277	14,392,277
R-squared	0.025	0.074	0.083	0.089

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (1), with Republicans as treatment group and Democrats as control group. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age).

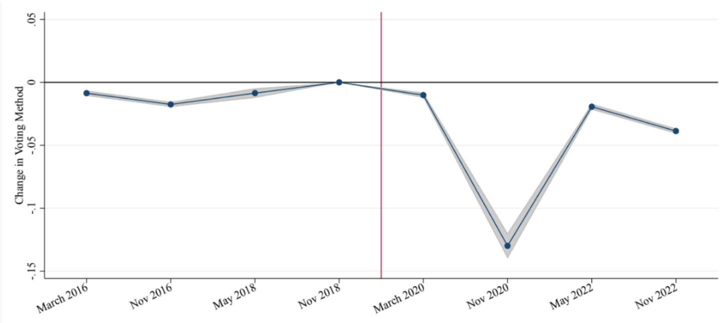
Table 6 displays paneled cross-sectional difference-in-differences results from the same dataset and are visually represented in Figure[4] .Republican voters were consistently less inclined to vote-by-mail following the post-treatment elections as compared to their Democratic counterparts, ranging from a 1.0 percentage point difference in the primary election of 2020, to a dramatic 13.0 percentage point difference in the general elections of the same year. It should be noted that VBM numbers spiked in the general election of 2020 at the height of the coronavirus pandemic, and this 13.0 percentage point decline (best visualized in Figure 4) is to be taken relative to the overall 19.4 – 19.6 percentage point increase in

4 For full table of estimation coefficients, see Appendix Table A1.

VBM, as seen in the linear estimator for *G20*. As discussed previously and represented in Figure 1, the COVID-19 pandemic presented a unique exogenous shock that inflated nontraditional voting methods' popularity, particularly VBM. While voters across all party lines elected to vote nontraditionally, Democratic voters who preferred voting by mail greatly outnumbered Republican voters. The persistence of this effect remains unclear, though estimators for both elections of 2022 signal a growing 3-percentage point difference in VBM popularity among Republican voters. Again, the interaction coefficients for certain pre-treatment covariates are statistically significant at the 5% level, prtentially signaling pre-treatment trends, likely due to the overwhelmingly large sample size (N = 14,392,277). These effects will be further examined in Section 5.4 – 5.5 as an additional robustness check.

These effects, in direct comparison to nontraditional voting (Table 4 & 5), carry over in direction though not in magnitude. Absentee voting spikes higher in the general election of 2020 as VBM, at about a 28-percentage point increase. This is to be expected, as nontraditional absentee voting encompasses early voting (including early curbside voting) and VBM. Despite this, the difference in absentee voting rates between Republican and Republican and Democratic voters is negligible in the general election of 2020.

Figure 4: Difference-In-Difference Estimates On VBM



Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (2), with Republicans as treatment group and Democrats as control group. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age).

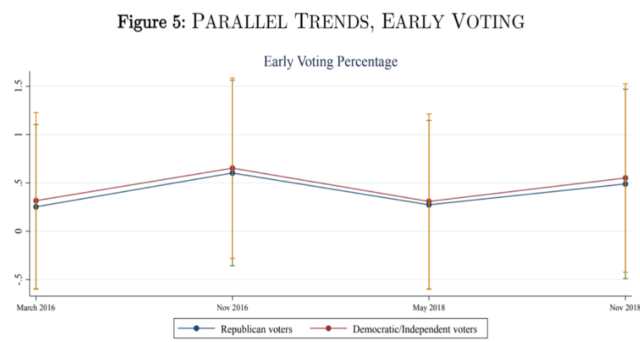
Table 6: Dyanmic Difference-In-Differences, Vote-By-Mail

Variables	(1)	(2)	(3)
REP	0.0115***	0.0107***	-0.00276***
	(0.00146)	(0.00139)	(0.000731)
REP × P16	-0.00934***	-0.00891***	-0.00861***
	(0.000957)	(0.00112)	(0.00106)
REP × G16	-0.0179***	-0.0173***	-0.0175***
	(0.000967)	(0.00101)	(0.00101)
REP × P18	-0.0130***	-0.00987***	-0.00855***
	(0.00124)	(0.00194)	(0.00193)
REP × P20	**-0.0132*** **	**-0.00971*** **	**-0.0102*** **
	(0.00108)	(0.00110)	(0.00106)
REP × G20	**-0.130*** **	**-0.130*** **	**-0.130*** **
	(0.00513)	(0.00511)	(0.00511)
REP × P22	-0.0218***	-0.0197***	-0.0194***
	(0.00117)	(0.00122)	(0.00117)
REP × G22	-0.0411***	-0.0399***	-0.0387***
	(0.00113)	(0.00108)	(0.00103)
P16	-0.0130***	-0.0130***	-0.0152***
	(0.000826)	(0.000878)	(0.000963)
G16	0.000647	0.00181**	0.00229**
	(0.000838)	(0.000880)	(0.000902)
P18	-0.0168***	-0.0159***	-0.0198***
	(0.000858)	(0.00123)	(0.00144)
P20	-0.0137***	-0.0152***	-0.0166***
	(0.000700)	(0.000718)	(0.000817)
G20	**0.194*** **	**0.195*** **	**0.196*** **
	(0.00716)	(0.00723)	(0.00724)
P22	-0.00352***	-0.00406***	-0.00676***
	(0.000782)	(0.000714)	(0.000785)
G22	0.0285***	0.0282***	0.0267***
	(0.00134)	(0.00130)	(0.00129)
Constant	0.0244***	0.0242***	-0.104***
	(0.000992)	(0.00157)	(0.00890)
Zip Code FE	NO	YES	YES
Demographic Controls	NO	NO	YES
Observations	14,392,277	14,392,277	14,392,277
R-squared	0.081	0.090	0.096

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (2), with Republicans as treatment group and Democrats as control group. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age).

Figure 5: Parallel Trends, Early Voting



5.3 Do Republicans voters vote early less?

Claims of fraudulent activity attributed to VBM via absentee ballots by key Republican leaders may have additional effects that spill over to another widely used nontraditional voting method: early voting. Figure 5 displays average rates of early voting and their respective 95% confidence intervals by Republican voters as compared to Democratic and Independent voters. As before, Democratic/Independent voters serve as a useful and reliable control group as they would have been exposed to nearly all other influencing factors that occurred in the examined period. The parallel trends assumption, by visible inspection, can be assumed to be true. Noticeably, Republican voters tend to vote early at visibly lower rates than their Democratic/Independent counterparts, with a widening gap in the general elections of 2018 ahead of the treatment periods. Table 6B depicts the coefficients of interest and confidence intervals from the Equation (1) on early voting.

Table 6: Pooled Difference-In-Differences, Early Voting

Variables	(1)	(2)	(3)	(4)
REP	**-0.0821*** **	**-0.0852*** **	**-0.0806*** **	**-0.0527*** **
	(0.00530)	(0.00601)	(0.00468)	(0.00680)
POST	**0.0164*** **	-	-	-
	(0.00354)			
POST × REP	**0.0581*** **	**0.0517*** **	**0.0521*** **	**0.0519*** **
	(0.00351)	(0.00334)	(0.00338)	(0.00323)
WHI				-0.00260
				(0.00276)
AAPI				-0.00867*
				(0.00505)
BLA				0.0694***
				(0.00406)
NAT				-0.0357***
				(0.00944)
MIX				0.0328***
				(0.00442)
HISP				-0.0160***
				(0.00229)
MALE				-0.0105***
				(0.000548)
LN_AGE				0.153***
				(0.00755)
Constant	0.548***	0.559***	0.557***	-0.0871**
	(0.00729)	(0.00670)	(0.00184)	(0.0340)
Election FE	NO	YES	YES	YES
Zip Code FE	NO	NO	YES	YES
Demographic Controls	NO	NO	NO	YES
Observations	14,392,277	14,392,277	14,392,277	14,392,277
R-squared	0.006	0.081	0.101	0.111

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (1), with Republicans as treatment group and Democrats as control group. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age).

Surprisingly, Republican voters were more likely to vote early post-treatment, by 5.2 – 5.8 percentage points. Estimators of the linear term, REP, indicated a negative bias for early voting among Republican voters, despite a positive effect during the post treatment periods. This effect ranged from a 5.2 – 8.5 percentage point decline among Republican voters as a whole, relative to Democratic voters. Following treatment, early voting rates rose by approximately 1.6% (*POST*) for all voters in the e timation. This effect, coupled with the consistently significant and positive ~5 percentage point difference in early voting rates for Republican voters in post-treatment periods, is especially notable. As before, age positively and directly correlates with the likelihood of voting early, with mixed results across racial demographic groups.

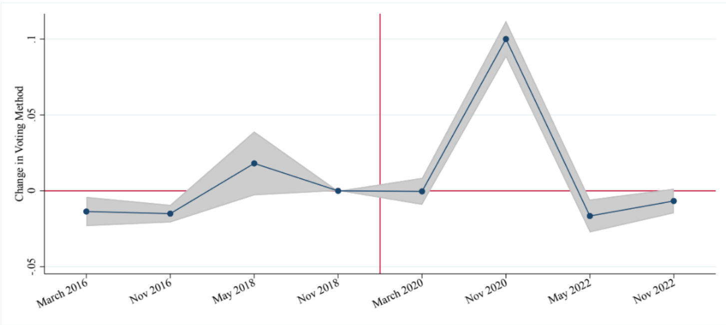
Table 7: Dynamic Difference-In-Differences, Early Voting

Variables	(1)	(2)	(3)
REP	-0.0500***	-0.0449***	-0.0317***
	(0.000512)	(0.000527)	(0.000539)
REP × P16	-0.0133***	-0.0124***	-0.0136***
	(0.000734)	(0.000748)	(0.000746)
REP × G16	-0.0116***	-0.0124***	-0.0150***
	(0.000614)	(0.000626)	(0.000624)
REP × P18	0.0147***	0.0156***	0.0181***
	(0.00109)	(0.00110)	(0.00110)
REP × P20	0.00273***	0.00316***	-0.000351
	(0.000821)	(0.000834)	(0.000830)
REP × G20	0.102***	0.102***	0.1000***
	(0.000598)	(0.000607)	(0.000607)
REP × P22	-0.0227***	-0.0194***	-0.0166***
	(0.000950)	(0.000948)	(0.000943)
REP × G22	-0.00581***	-0.00622***	-0.00668***
	(0.000667)	(0.000668)	(0.000666)
P16	-0.235***	-0.236***	-0.244***
	(0.000458)	(0.000466)	(0.000465)
G16	0.101***	0.102***	0.103***
	(0.000353)	(0.000360)	(0.000359)
P18	-0.242***	-0.242***	-0.258***
	(0.000621)	(0.000627)	(0.000626)
P20	-0.175***	-0.176***	-0.181***
	(0.000444)	(0.000452)	(0.000451)
G20	0.102***	0.108***	0.112***
	(0.000399)	(0.000406)	(0.000405)
P22	-0.127***	-0.128***	-0.143***
	(0.000588)	(0.000586)	(0.000581)
G22	0.0206***	0.0222***	0.0203***
	(0.000393)	(0.000395)	(0.000393)
Constant	0.551***	0.548***	-0.0627***
	(0.000324)	(0.000330)	(0.00300)
Zip Code FE	NO	YES	YES
Demographic Controls	NO	NO	YES
Observations	21,158,464	20,497,852	20,497,852
R-squared	0.073	0.093	0.105

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (2), with Republicans as treatment group and Democrats and Independents as control group. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age).

Figure 6: Difference-In-Differences Estimates On Early Voting



Estimates from Equation (2) for early voting are displayed in Table 7 and plotted in Figure 6[5]. Notably, Republican voters were significantly more inclined to vote early as compared to Democratic/Independent voters in the general election of 2020 by about 10 percentage points, a dramatic inverse of the effect on VBM as displayed in Figure 4 and Table 6. However, in the primary and general elections of 2022, this effect is immediately negated, though differences in average early voting rates are relatively small by party affiliation. Estimates on linear racial demographic var ables vary widely between groups, though lack significance even at the 10% significance level. With the exception of the general election in 2018, Republican voters display negligible changes in early voting behavior relative to Democratic or Independent voters. To better understand the mechanisms at work, it may be prudent to examine these effects when aggregated by race or political party strength, as in Section 5.4 – 5.5. Doing so also reduces the sample sizes of these estimations and achieves more statistically meaningful significance.

5.4 Voting Method by Race

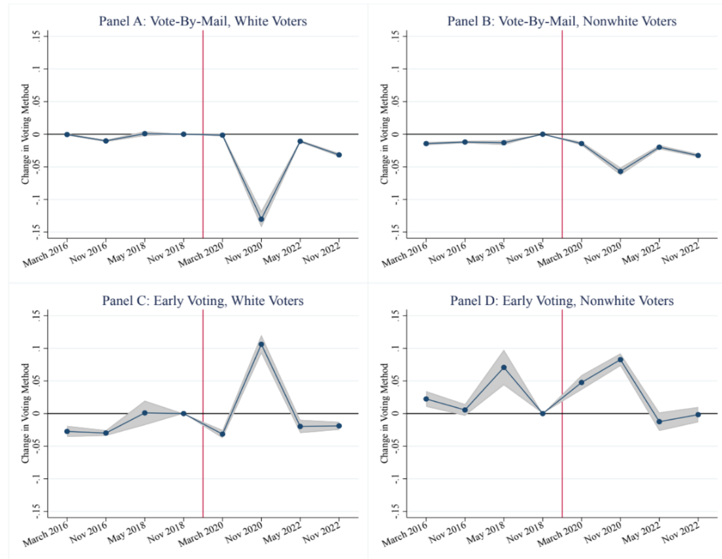
Applying the same DD model in Equation (2), the effects of fraud alleg tions on VBM and absentee voting can also be assessed on voting behavior by racial group:

$$y_{i,t} = \beta_0 + \sum_{j=1}^{8} \beta_j \cdot Party_i \cdot Election_{j=t} + \sum_{j=9}^{16} \beta_j \cdot Election_{j=t+8} + \beta_k \cdot Party_i + \beta_m \cdot Z_{i,t} + \delta_m + \varepsilon_{i,t}, \qquad \text{where } k \neq m \neq j. \tag{3}$$

5 For full table of estimation coefficients, see Appendix Table A1.

where $y_{i,t}$ represents voting method utilized among either white voters or nonwhite voters for individual i at election t, and $Z_{i,t}$ only accounts for gender, age, and geographical control variables. Figure 7 plots the seprate interaction estimates of β_j from equation (2) for white voters and nonwhite voters, with standard errors clustered at the city level[6].

Figure 7: Effect On Voting Method By Race



Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (2). Standard errors are clustered at the city level and 95% confidence intervals are shaded in gray. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Hispanic/Latino, Age).

These estimates complement the county-level results presented earlier. Republican voters, as before, are less inclined to vote by mail on and after the 2020 elections, particularly in the general election of 2020. Notably, white, Republican voters are significantly less likely to use VBM as compared to white, Democratic voters. This difference is less distinct for nonwhite Republican voters versus nonwhite Democratic voters. The same patterns, though inversed, are reflected in early voting analysis.

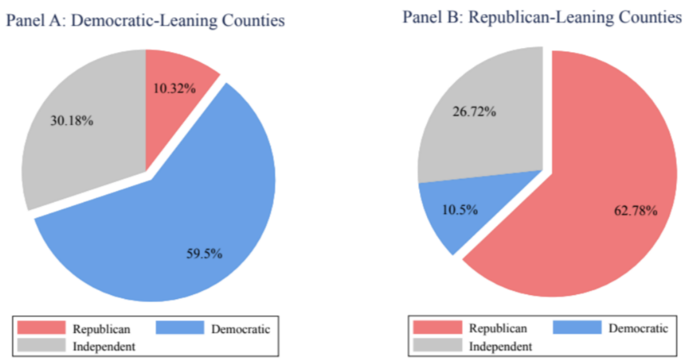
It should be noted that pre-treatment coefficients for VBM and early voting estimations are statistically significant, potentially signaling a violation of the parallel trends' assumption, particularly among nonwhite Republican voters relative to nonwhite Democrat voters. The large sample

6 See Appendix Table A3 and Table A4 for full table of estimation coefficients.

size (over 4 million voters) may contribute to smaller standard errors and thus, higher likelihood for statistical significance. As a robustness check, Section 5.5 explores the same effects on smaller, county-wide sa ples by political party strength, followed by an additional difference-in-differences analysis with a "placebo" test in Section 5.6

5.5 Voting Method by Party Strength

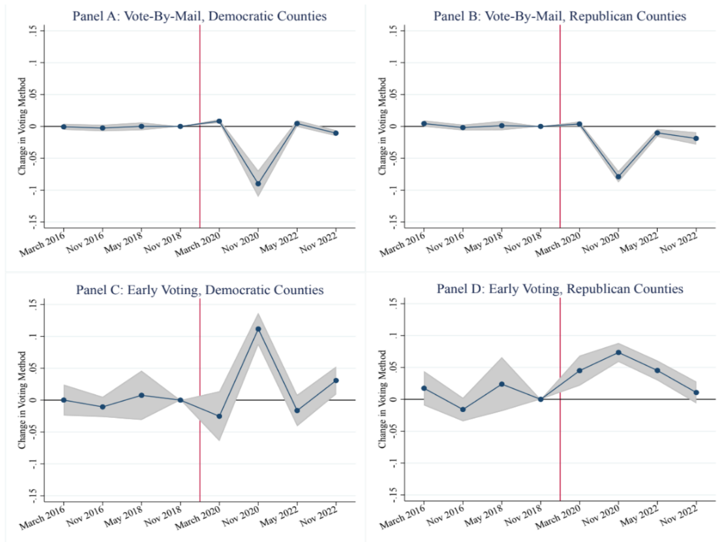
Figure 8: Countywide Party Affiliation



As a robustness check, a county-level analogue of Equation (2) can assess exaggerated impacts of the misinformation campaign launched in 2020 by Republican leaders on various counties in North Carolina by political party strength. To satisfy the necessary parallel trends assumption necessitated by application of DD estimation and to better understand the extent of the examined effects on counties dominated by a single political party, voter registration and voter history data from three North Carolinian counties with the highest registered Democrat-to-Republican ratio, as well as with the highest Republican-to-Democrat ratio. Using data from the North Carolina Board of Elections, party affiliation breakdown from heavily democratic counties (Hertford County, Northampton County, Durham County) and heavily republican counties (Mitchell County, Avery County, Yadkin County) are displayed in Figure 8.

Coefficients and confidence intervals from estimation of the subsequent DD model (Equation (2)) are plotted in Figure 9, with estimation coefficients displayed in Appendix Table A2. In contrast to prior estimations of statewide voter behavior, there is little evidence of differential trends between Democractic/Independent voters (control groups) and Republican voters (treatment group) prior to the allegations of voter fraud in 2020, satisfaying the parallel trends assumption. All pretreatment estimates are less than 2.5 percentage points in magnitude and are insignificantly different from 0.

Figure 9: Effect On Voting Method By Party Strength



Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (2). Standard errors are clustered per zip code and 95% confidence intervals are shaded in gray. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Hispanic/Latino, Age).

Post-treatment, Republican voters in both Republican-dominated and Democratic dominated counties display similar VBM behavior as seen in the statewide analysis, in Table 4. Relative to Democratic/Independent voters, Republican voters are less likely to vote by mail following the 2020 elections, with a particularly large shock in the general election of 2020: 9 percentage point decline in VBM rates in heavily Democratic counties compared to a 7.9 percentage point difference in heavily Republican counties. However, these rates normalize and bounce back much quicker in Democratic-majority counties in the 2022 elections, while Republican-majority counties display a persistent significant difference in VBM rates between Democratic/Independents and Republican voters.

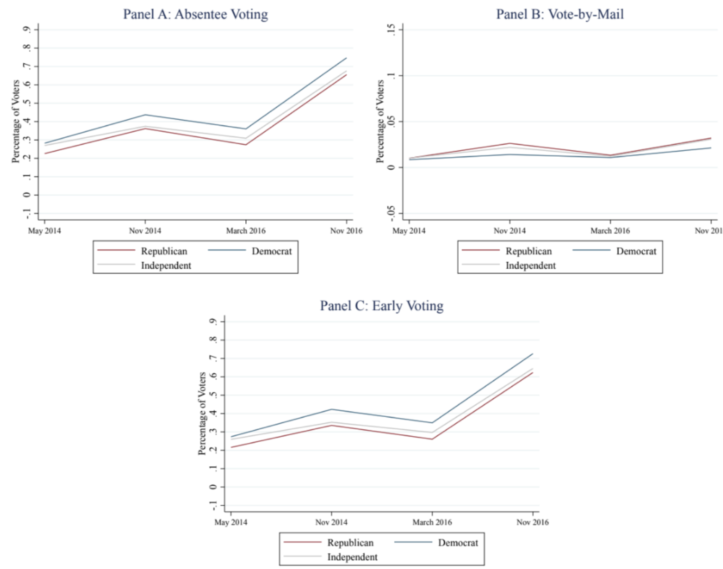
As for early voting, Republican voters in Democratic-majority counties display vastly different behavior as compared to Republican-dominated counties. In the primary election of 2020 (the first posttreatment election), Republican voters in Republican majority counties are actually *more* likely to vote early relative to Democratic/Independent voters by a significant 4.5 percentage points, in contrast to a negative effect of 2.5 percentage points in Democratic-majority counties. Similar to VBM estimators, the shock due to COVID-19 in the general elections of 2020 led to a more dramatic effect for Republican voters in Democratic-majority counties (11.2 percentage point increase) as compared to Republican-majority counties (7.4 percentage point increase), with this effect bouncing back quickly to 0 in the former and remaining persistently positive in the latter counties. The majority of post-treatment estimators are statistically significant at the 95% confidence interval in both estimations.

The findings displayed in Figure 9 are surprising and shed further light on the preliminary results from statewide analyses in Tables 5 – 8. In counties where most registered voters are Democratic, Republican voters respond more intensely to both treatment and shocks. In Republican-majority counties, Republican voters respond less dramatically relative to Democratic/Independent voters, though effects are far more persistent.

5.6 Additional Robustness Checks

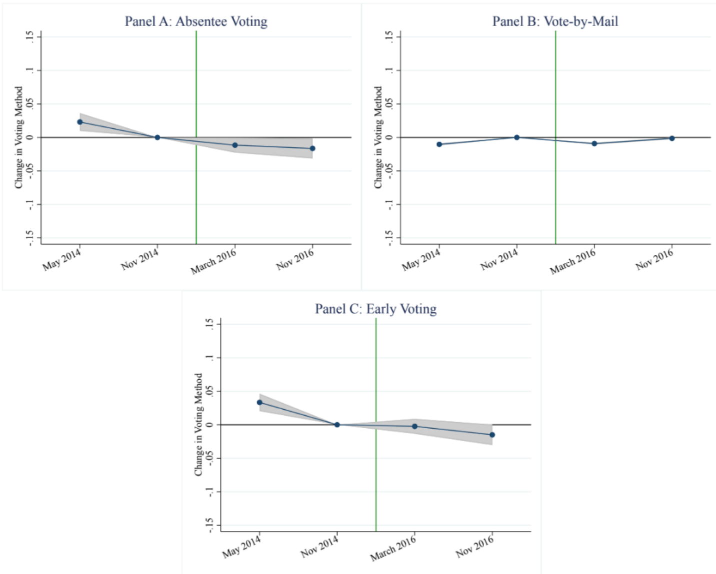
Literature in the field of policy analysis and political economy often rely on additional strategies to test a dynamic difference-in-differences model's robustness to check for equality of pre-treatment trends. One such reliable strategy constructs a false treatment as a placebo effect, usually performed prior to actual treatment year (Hanna & Oliva, 2015; Slusky, 2017; Baker et al. 2019). In a similar fashion, statewide voter-level voter registration and voter history data from the primary and general elections 2014 to 2016 were compiled and aggregated following the same methods as described in Chapter 3. The earliest election is limited to May 2014, as the North Carolina State Board of Elections does not maintain individual-level voter history data prior to 2014. As there were no significant voter legislation enacted during this period and allegations of voter fraud via absentee ballots or VBM, one wouldn't reasonably expect significant variations in voting methods between Republican voters and Democratic voters. Figure 10 displays the raw rends of average rates of each voting method during 2014 to 2016, by political party. Clearly, early voting and absentee voting rates between Republican voters and Democratic voters closely mirror one another, likely due to low rates in VBM during this period.

Figure 10: Voting Method Over Time (2014-2016)



Simulating a fake "placebo" treatment between the general election of 2014 and the primary election of 2016, the results for the dynamic difference-in-differences estimator in Equation (2) are plotted in Figure 11 and in Appendix Table 6. As expected, Republican voters' pre-"treatment" and post-"treatment" display little to no significant variation in absentee voting, VBM, and early voting rates relative to Democratic voters, with nearly all post-"treatment" estimators insignificant at the 95th confidence interval with the exception of VBM. The lower confidence interval bands displayed in Panel B for VBM rates is likely attributed to a much smaller sample size, as only 1 – 2% of all voters in each election elected to vote-by-mail.

Figure 11: Dynamic D.D. Estimates With Placebo Treatment



Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (2) with Republicans as treatment group and Democrats and Independents as control group. Standard errors are clustered at the city level and 95% confidence intervals are shaded in gray. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Hispanic/Latino, Age).

The results displayed in Figure 11 and Appendix Table A5, when in conjunction to the analysis conducted and results of Section 5.5, further corroborate the findings of this paper. Using individual-level data over time results in extremely large observational samples, which will nearly always result in significant results (Hill and Leamer 1980). The statistically insignificant estimators in the model with a fake, placebo treatment (Figure 11) and pre-treatment estimators from Figure 9 further validate and support the satisfaction of the parallel trends assumption.

6 Discussion

The following section provides plausible explanations for the findings from Chapter 5. While they are founded on the estimates from this paper's regvressions and supported by existing research, they are merely suggestions for future exploration and have not been conclusively proven.

6.1 Plausible Explanation of Results

The results from Chapter 5 support initial hypotheses that Republican voters significantly voted less by mail beginning in the 2020 primary and general elections, following the widespread misinformation campaign by GOP officials and leaders against VBM and absentee voting. This effect is particularly significant and persistent for Republican voters in Republican-dominated counties, while the magnitude of the spike in VBM numbers in the general election of 2020 was higher for Republican voters in Democratic-dominated counties. This may be attributed to environmental factors—a Republican voter in a Democratic-majority county may feel pressured to conform to popular voting methods such as early voting or VBM on low-stakes elections. In contrast, as the general election of 2020 also happened to be a presidential election, Republican voters in heavily Democratic counties may have been more influenced by the voter fraud narrative surrounding VBM and chose to move away from more convenient methods of voting, even at the peak of the coronavirus pandemic. In heavily Republican counties, Republican voters as a whole are likely more concerned with electoral integrity and persistently choose to avoid vote-by-mail relative to Democratic/Independent voters, with this negative trend continuing years after the peak of the pandemic and the height of electoral fraud claims.

With regards to early voting, the statewide cross-sectional DD logistic regression model displays conflicting results, with a positive coefficient on the presidential election of 2020, a negative coefficient on the primary election of 2022, and statistically insignificant results on all other post-treatment elections. However, as displayed in Figure 9, estimates on the early voting rates by party strength shed further light on the complex dynamics at play. Republican voters in heavily Democratic counties are less likely to vote early relative to Republican voters in heavily Republican counties, with only statistically positive effects on two post-treatment elections. In the latter group, Republican voters were more likely to vote early across nearly all post-treatment elections. These findings are surprising and go against initial hypotheses that claims of electoral fraud would discourage Republican voters from voting nontraditionally, including voting early. Since voting early in North Carolina requires an absentee ballot and Republican officials have repeatedly raised concerns regarding missing or miscounted absentee ballots, it is surprising that Republican voters are more likely than Democratic or Independent voters to vote early. This phenomenon can be plausibly explained by the significantly negative effect of Republican voters using VBM. Since the

COVID-19 pandemic raised safety concerns of voting in-person on Election Day, Republican voters could have likely turned to the only other alternative voting method: early voting. In this case, the increase in early voting among Republican voters' post treatment can be seen as a conscious choice between the lesser of two evils.

Further, as seen in Figure 7, white Republican voters are significantly more likely to VBM and less likely to vote early relative to nonwhite Republican voters. However, due to the significance of pre-treatment coefficients potentially signaling the violation of the DD parallel trends assumption, this findings cannot be concluded with reasonable certainty.

6.2 Key Implications

The implications of these findings are important in the context of voter restriction legislation targeting VBM or access to voting in general. While existing research has not linked the rise of VBM to any significant increase in a political party's turnout or vote share in an election, these studies are based in pre-pandemic times and before the validity of VBM and absentee voting was questioned and targeted by GOP officials (Thompson et al 2020, Barber and Holbein 2020, Elul et al 2017). Assuming the persistence of the effects described above, future legislation backed by GOP officials restricting VBM may begin to disproportionately affect Democratic and Independent voters. Universal VBM has already been shown to increase voter turnout among minority groups and younger voters, demographic groups that increasingly identify as Democratic (Atsusaka et al 2019, Griffin 2021).

Attempts to restrict voting access across the United States have exploded in recent years. According to the Brennan Center of Justice, Republicans have introduced, filed, or passed over 250 bills across 48 states to restrict voting access as of May 2021 (Berry et al., 2021). 125 of these 250 bills specifically target vote-by-mail or include provisions that further restrict VBM. In 2017, the North Carolina state legislature introduced Senate Bill 824, requiring voter ID for in-person voting and tightening restrictions on alternative forms of voting (S.B. 824, 2017). While the law was ultimately struck down by the state's Supreme Court in violation of the North Carolina Constitution's Equal Protection clause, legislation such as these cannot be overlooked as attempts to discourage or otherwise disenfranchise eligible voters.

6.3 Limitations and Opportunities for Future Research

The central limitation this paper faces is the assumption that the COVID-19 pandemic was an exogenous shock that influenced the voting behavior of all voters equally, regardless of political party affiliation. Republican voters may have been affected by the pandemic in various ways that the control variables in the cross-sectional difference-in differences model could not account for. Perhaps the quicker spread of COVID-19 within urban areas with higher population density influenced the voting behavior of Democratic voters, or perhaps certain voters are more risk-averse and health-conscious and therefore, were more likely to vote nontraditionally. These variables are difficult to account for, though demographic and geographic controls were implemented throughout all regressions.

In many ways, North Carolina was an ideal choice to conduct this paper's research, as the state did not experience major changes in voting policy between the examined years of 2016 to 2022, and its status as a no-excuse absentee voting state provides a wide range of methods for voters to drop their ballots. As a battleground state, North Carolina is home to a wide range of counties with varying political party dominance and the state's Board of Elections regularly maintains a comprehensive voter history and registration database. Despite this, the cross-sectional logit difference-in differences model applied in this paper requires no pre-treatment trends between Democratic/Independent voters and Republican voters. While the estimators on the effect of perceived electoral fraud on counties with a dominant political party certainly met this criteria as shown in Figure 9 (Section 5.5), the statewide estimators sometimes had pre-treatment trends that may potentially confound the results. These limitations, however, were addressed in Section 5.6 with additional robustness checks.

Ideally, future research in this area should examine the persistence and magnitude of the effects in the upcoming elections of 2024 and 2026. Incorporating additional post-treatment years can shed further light on the mechanisms behind the effects of perceived electoral insecurity on voting behavior in the long run.

7 Conclusion

This paper explores the effect of perceived electoral insecurity amidst claims of voter fraud on voting behavior using individual-level voter data preceding and following the 2020 U.S. election cycle. Since alleged fraud via mail-in voting is a relatively novel issue that only recently gained extensive media coverage, its real impacts on voting behavior have not been comprehensively studied. Utilizing a series of cross-sectional difference-in differences analyses and further supported by robustness checks via a placebo treatment analysis, we find that Republican voters were significantly less likely to vote by mail and more likely to vote early immediately following the 2020 primary and general elections as compared to Democratic and Independent voters.

To reduce noise and better understand the mechanisms behind the statewide estimation, analysis performed on white Republican voters versus nonwhite Republican voters as well as Republican-majority counties versus Democratic-majority counties reveals further insights. The effects aforementioned are more persistent among nonwhite Republican voters, as a greater number of nonwhite Republicans chose not to vote by mail relative to nonwhite Democrats compared to white Republicans and white Democrats. These effects, however, are greater in magnitude for white Republicans relative to white Democrats, particularly in the general election of 2020.

The same results apply to Republican-majority counties relative to Democratic majority counties, with greater significance and robustness. Republicans in Republican majority counties were similarly more consistent in rejecting VBM and opting instead for early absentee voting for several post-treatment periods, as compared to Republicans in Democratic-majority counties. As hypothesized in Chapter 6, this may be attributed to the social pressure or influence of a largely Democratic county on Republican voters, who may feel pressured to conform to popular voting methods such as early voting or VBM on low-stakes elections, thus explaining the relatively lower effects as seen in the presidential election in 2020. In heavily Republican counties, Republican voters as a whole are likely more concerned with electoral integrity and persistently choose to avoid vote-by-mail relative to Democratic/Independent voters, with this negative trend continuing years after the peak of the pandemic and the height of electoral fraud claims.

The electoral misinformation campaign waged against VBM and absen-

tee voting, coupled with the COVID-19 pandemic, led to historically unprecedented and precarious elections in 2020 and beyond. This paper underscores the importance of maintaining electoral integrity and highlights the consequences and mechanisms at work when the fundamental right to vote is under attack.

8 Appendix

Appendix Table A1: Dynamic Difference-In-Differences

VARIABLES	(1) Absentee Voting	(2) Vote-By-Mail	(3) Early Voting
REP	-0.0566***	-0.00276***	-0.0538***
	(0.00903)	(0.000731)	(0.00871)
REP_P16	-0.00822	-0.00861***	0.000388
	(0.00576)	(0.00106)	(0.00596)
REP_G16	-0.0264***	-0.0175***	-0.00887**
	(0.00358)	(0.00101)	(0.00382)
REP_P18	0.0391***	-0.00855***	0.0476***
	(0.0135)	(0.00193)	(0.0123)
REP_P20	0.0107**	-0.0102***	0.0209***
	(0.00524)	(0.00106)	(0.00551)
REP_G20	-0.000393	-0.130***	0.129***
	(0.00388)	(0.00511)	(0.00720)
REP_P22	-0.0263***	-0.0194***	-0.00691
	(0.00692)	(0.00117)	(0.00691)
REP_G22	-0.0421***	-0.0387***	-0.00340
	(0.00510)	(0.00103)	(0.00530)
P16	-0.267***	-0.0152***	-0.252***
	(0.00606)	(0.000963)	(0.00538)
G16	0.111***	0.00229**	0.109***
	(0.00427)	(0.000902)	(0.00449)
P18	-0.300***	-0.0198***	-0.280***
	(0.0123)	(0.00144)	(0.0111)
P20	-0.213***	-0.0166***	-0.196***
	(0.00479)	(0.000817)	(0.00433)
G20	0.284***	0.196***	0.0887***
	(0.00658)	(0.00724)	(0.00911)
P22	-0.153***	-0.00676***	-0.146***
	(0.00809)	(0.000785)	(0.00788)
G22	0.0498***	0.0267***	0.0231***
	(0.00341)	(0.00129)	(0.00382)
WHI	-0.0189***	-0.0158***	-0.00313
	(0.00267)	(0.00229)	(0.00276)
AAPI	0.0151***	0.0234***	-0.00833*
	(0.00514)	(0.00276)	(0.00501)
BLA	0.0176***	-0.0518***	0.0694***
	(0.00531)	(0.00213)	(0.00400)
NAT	-0.0671***	-0.0315***	-0.0356***
	(0.00976)	(0.00234)	(0.00939)
MIX	0.0247***	-0.00842***	0.0332***
	(0.00372)	(0.00181)	(0.00441)
HISP	-0.0328***	-0.0184***	-0.0144***
	(0.00298)	(0.00177)	(0.00228)
MALE	-0.0157***	-0.00529***	-0.0104***
	(0.000634)	(0.000253)	(0.000545)
LN_AGE	0.193***	0.0398***	0.153***
	(0.00743)	(0.00200)	(0.00750)
Constant	-0.167***	-0.104***	-0.0630*
	(0.0333)	(0.00890)	(0.0358)
Observations	14,392,277	14,392,277	14,392,277
R-squared	0.177	0.096	0.113
Zip Code FE	YES	YES	YES
Demographic Controls	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (2). Treatment group: Republican voters. Control group: Democratic voters. While demographic variables were omitted in Table 4, Table 6, and Table 8, they are on display here. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age).

Appendix Table A2: Dynamic Difference-In-Differences By Party Strength

	Democratic-majority Counties		Republic-majority Counties	
	(1)	(2)	(1)	(2)
Variables	VBM	Early Voting	VBM	Early Voting
REP_P16	-0.000654	9.99e-05	0.00445*	0.0174
	(0.00229)	(0.0120)	(0.00231)	(0.0130)
REP_G16	-0.00252	-0.0106	-0.00179	-0.0160*
	(0.00245)	(0.00781)	(0.00218)	(0.00891)
REP_P18	7.66e-05	0.00757	0.00117	0.0238
	(0.00300)	(0.0191)	(0.00345)	(0.0205)
REP_P20	0.00824***	-0.0253	0.00379*	0.0450***
	(0.00145)	(0.0193)	(0.00186)	(0.0114)
REP_G20	-0.0900***	0.112***	-0.0789***	0.0735***
	(0.0102)	(0.0125)	(0.00433)	(0.00718)
REP_P22	0.00456*	-0.0164	-0.0101***	0.0453***
	(0.00254)	(0.0122)	(0.00292)	(0.00754)
REP_G22	-0.0105***	0.0308***	-0.0189***	0.0107
	(0.00220)	(0.0108)	(0.00463)	(0.00827)
P16	-0.00426*	-0.253***	-0.00454	-0.212***
	(0.00244)	(0.00924)	(0.00301)	(0.0189)
G16	0.0109***	0.158***	0.00631**	0.105***
	(0.000834)	(0.00794)	(0.00302)	(0.0138)
P18	-0.00463	-0.353***	-0.00568*	-0.240***
	(0.00340)	(0.0224)	(0.00282)	(0.0190)
P20	-0.00988***	-0.207***	-0.00263	-0.168***
	(0.00110)	(0.00862)	(0.00223)	(0.0112)
G20	0.229***	0.0718***	0.127***	0.192***
	(0.0156)	(0.0234)	(0.00531)	(0.0213)
P22	0.00162*	-0.203***	0.0115***	-0.133***
	(0.000901)	(0.0121)	(0.00389)	(0.0187)
G22	0.0297***	0.00903	0.0217***	-0.00564
	(0.00209)	(0.00715)	(0.00459)	(0.00837)
Constant	0.0202***	0.613***	0.0163***	0.387***
	(0.00292)	(0.00915)	(0.00246)	(0.0132)
Observations	779,715	779,715	165,071	165,071
R-squared	0.445	0.432	0.473	0.536
Zip Code FE	YES	YES	YES	YES
NCID FE	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (2). Standard errors are clustered per zip code. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Hispanic/Latino, Age).

Appendix Table A3: Dynamic Difference-In-Differences, Vote-By-Mail By Race

	White Voters		Nonwhite Voters	
Variables	(1)	(2)	(3)	(4)
REP	0.00636***	0.00597***	0.0147***	-0.00674***
	(0.000747)	(0.000688)	(0.00132)	(0.00133)
REP_P16	4.61e-05	-0.000477	-0.0146***	-0.0154***
	(0.000717)	(0.000719)	(0.00121)	(0.00117)
REP_G16	-0.00959***	-0.0102***	-0.0117***	-0.0120***
	(0.000762)	(0.000765)	(0.000893)	(0.000876)
REP_P18	0.000721	0.000908	-0.0144***	-0.0123***
	(0.00170)	(0.00178)	(0.00178)	(0.00182)
REP_P20	-9.64e-05	-0.00143	-0.0142***	-0.0146***
	(0.00104)	(0.000988)	(0.00133)	(0.00133)
REP_G20	-0.129***	-0.129***	-0.0567***	-0.0566***
	(0.00638)	(0.00640)	(0.00301)	(0.00297)
REP_P22	-0.0107***	-0.0107***	-0.0208***	-0.0213***
	(0.000857)	(0.000832)	(0.00161)	(0.00158)
REP_G22	-0.0313***	-0.0315***	-0.0330***	-0.0332***
	(0.00127)	(0.00131)	(0.00125)	(0.00123)
P16	-0.0138***	-0.0156***	-0.0101***	-0.0107***
	(0.00106)	(0.00113)	(0.000707)	(0.000702)
G16	0.00992***	0.0101***	-0.000540	-0.000241
	(0.00115)	(0.00110)	(0.000577)	(0.000562)
P18	-0.0172***	-0.0203***	-0.00982***	-0.0111***
	(0.00130)	(0.00148)	(0.00105)	(0.00116)
P20	-0.0163***	-0.0171***	-0.0129***	-0.0133***
	(0.00108)	(0.00109)	(0.000450)	(0.000449)
G20	0.200***	0.201***	0.151***	0.152***
	(0.0104)	(0.0104)	(0.00563)	(0.00558)
P22	-0.00456***	-0.00728***	-0.00520***	-0.00660***
	(0.000783)	(0.000814)	(0.000632)	(0.000652)
G22	0.0275***	0.0272***	0.0236***	0.0224***
	(0.00177)	(0.00176)	(0.00129)	(0.00130)
HISP		0.00542**		-0.0280***
		(0.00248)		(0.00179)
MALE		-0.00770***		-0.00211***
		(0.000418)		(0.000374)
LN_AGE		0.0381***		0.0332***
		(0.00177)		(0.00217)
AAPI				0.0236***
				(0.00280)
BLA				-0.0467***
				(0.00224)
NAT				-0.0337***
				(0.00234)
MIX				-0.00517***
				(0.00184)
Constant	0.0241***	-0.126***	0.0200***	-0.0743***
	(0.00202)	(0.00729)	(0.00138)	(0.00882)
Observations	15,001,823	15,001,823	5,496,029	5,496,029
R-squared	0.094	0.097	0.087	0.095
Zip Code FE	YES	YES	YES	YES
Demographic Controls	NO	YES	NO	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Estimation of coefficients as pictured in Figure 7. Data from NCSBE. DD coefficients are from the estimation of Equation (2). Standard errors are clustered at the city level. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Hispanic/Latino, Age).

Appendix Table A4: Dynamic Difference-In-Differences, Early Voting by Race

	White Voters		Nonwhite Voters	
Variables	(1)	(2)	(3)	(4)
REP	-0.0187***	-0.0212***	-0.103***	-0.0685***
	(0.00518)	(0.00558)	(0.00558)	(0.00687)
REP_P16	-0.0256***	-0.0277***	0.0205***	0.0244***
	(0.00416)	(0.00430)	(0.00595)	(0.00598)
REP_G16	-0.0271***	-0.0296***	0.00764*	0.00553
	(0.00210)	(0.00215)	(0.00459)	(0.00452)
REP_P18	1.06e-05	0.000441	0.0613***	0.0724***
	(0.00926)	(0.00952)	(0.0135)	(0.0140)
REP_P20	-0.0263***	-0.0318***	0.0488***	0.0488***
	(0.00328)	(0.00319)	(0.00587)	(0.00564)
REP_G20	0.107***	0.106***	0.0837***	0.0819***
	(0.00729)	(0.00726)	(0.00492)	(0.00486)
REP_P22	-0.0202***	-0.0203***	-0.0194***	-0.00981
	(0.00526)	(0.00515)	(0.00728)	(0.00714)
REP_G22	-0.0177***	-0.0186***	-0.00453	-0.000562
	(0.00319)	(0.00309)	(0.00611)	(0.00588)
P16	-0.239***	-0.246***	-0.230***	-0.244***
	(0.00509)	(0.00499)	(0.00582)	(0.00577)
G16	0.0874***	0.0882***	0.124***	0.127***
	(0.00360)	(0.00373)	(0.00512)	(0.00509)
P18	-0.244***	-0.257***	-0.238***	-0.263***
	(0.0107)	(0.0107)	(0.0119)	(0.0125)
P20	-0.162***	-0.165***	-0.199***	-0.210***
	(0.00387)	(0.00372)	(0.00541)	(0.00541)
G20	0.0886***	0.0925***	0.136***	0.143***
	(0.00926)	(0.00948)	(0.00646)	(0.00655)
P22	-0.143***	-0.154***	-0.101***	-0.126***
	(0.00765)	(0.00740)	(0.00834)	(0.00839)
G22	0.0195***	0.0181***	0.0295***	0.0222***
	(0.00403)	(0.00405)	(0.00430)	(0.00419)
HISP		0.0110***		-0.0192***
		(0.00311)		(0.00231)
MALE		-0.00821***		-0.0135***
		(0.000530)		(0.00121)
LN_AGE		0.151***		0.176***
		(0.00916)		(0.00672)
AAPI				-0.00866*
				(0.00522)
BLA				0.0668***
				(0.00250)
NAT				-0.0362***
				(0.0123)
MIX				0.0356***
				(0.00359)
Constant	0.536***	-0.0696*	0.569***	-0.174***
	(0.00493)	(0.0406)	(0.00351)	(0.0289)
Observations	15,001,823	15,001,823	5,496,029	5,496,029
R-squared	0.096	0.104	0.090	0.107
Zip Code FE	YES	YES	YES	YES
Demographic Controls	NO	YES	NO	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Estimation of coefficients as pictured in Figure 7. Data from NCSBE. DD coefficients are from the estimation of Equation (2). Standard errors are clustered at the city level. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Hispanic/Latino, Age).

Appendix Table A5: Dynamic Difference-In-Differences, Placebo Effect

	Absentee Voting		VBM		Early Voting	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
REP	-0.0440***	-0.0391***	0.00679***	0.00827***	-0.0508***	-0.0474***
	(0.00653)	(0.00407)	(0.000720)	(0.000692)	(0.00651)	(0.00417)
REP_P14	0.0227***	0.0231***	-0.0100***	-0.0103***	0.0328***	0.0334***
	(0.00701)	(0.00673)	(0.000718)	(0.000805)	(0.00695)	(0.00660)
REP_P16	-0.0130**	-0.0115**	-0.00931***	-0.00913***	-0.00367	-0.00236
	(0.00597)	(0.00566)	(0.000606)	(0.000620)	(0.00598)	(0.00563)
REP_G16	-0.0163**	-0.0164**	-0.00138**	-0.00137**	-0.0149*	-0.0150*
	(0.00765)	(0.00764)	(0.000569)	(0.000571)	(0.00778)	(0.00777)
P14	-0.173***	-0.172***	-0.00667***	-0.00628***	-0.167***	-0.166***
	(0.00686)	(0.00690)	(0.000567)	(0.000507)	(0.00663)	(0.00661)
P16	-0.0805***	-0.0814***	-0.00383***	-0.00407***	-0.0767***	-0.0773***
	(0.00696)	(0.00654)	(0.000568)	(0.000578)	(0.00687)	(0.00644)
G16	0.310***	0.311***	0.00715***	0.00717***	0.303***	0.303***
	(0.00838)	(0.00835)	(0.000828)	(0.000831)	(0.00802)	(0.00800)
WHI	-0.0190***	-0.0205***	-0.00417***	-0.00331***	-0.0148***	-0.0172***
	(0.00265)	(0.00254)	(0.00106)	(0.00113)	(0.00295)	(0.00321)
AAPI	0.00294	0.00260	0.0133***	0.0102***	-0.0104	-0.00761
	(0.0103)	(0.00739)	(0.00233)	(0.00261)	(0.00998)	(0.00755)
BLA	0.0467***	0.0574***	-0.0160***	-0.0135***	0.0627***	0.0709***
	(0.00816)	(0.00588)	(0.000884)	(0.00107)	(0.00801)	(0.00531)
NAT	-0.126***	-0.0521***	-0.0151***	-0.00854***	-0.111***	-0.0436***
	(0.0189)	(0.0105)	(0.00124)	(0.00199)	(0.0185)	(0.0113)
MIX	0.0379***	0.0382***	-0.000584	-0.000437	0.0385***	0.0386***
	(0.00569)	(0.00587)	(0.00176)	(0.00171)	(0.00569)	(0.00566)
HISP	-0.0152***	-0.0148***	-0.00207	-0.00280*	-0.0132***	-0.0120***
	(0.00394)	(0.00322)	(0.00127)	(0.00155)	(0.00356)	(0.00287)
MALE	-0.0146***	-0.0135***	-0.00328***	-0.00320***	-0.0113***	-0.0103***
	(0.000886)	(0.000910)	(0.000256)	(0.000244)	(0.000777)	(0.000796)
LN_AGE	0.244***	0.241***	0.0104***	0.0115***	0.234***	0.230***
	(0.00717)	(0.00774)	(0.00160)	(0.00161)	(0.00644)	(0.00724)
Constant	-0.573***	-0.566***	-0.0176***	-0.0241***	-0.556***	-0.542***
	(0.0291)	(0.0336)	(0.00666)	(0.00704)	(0.0254)	(0.0322)
Observations	5,362,268	5,362,268	5,362,268	5,362,268	5,362,268	5,362,268
R-squared	0.146	0.169	0.005	0.010	0.139	0.164
Zip Code FE	NO	YES	NO	YES	NO	YES
Demographic Controls	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Estimation of coefficients as pictured in Figure 11. Data from NCSBE. DD coefficients are from the estimation of Equation (2). Standard errors are clustered at the city level. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Hispanic/Latino, Age).

Appendix Table A6: Pooled Difference-In-Differences, Absentee Voting (Including Independent Voters)

Variables	(1)	(2)	(3)	(4)
REP	-0.0557***	-0.0548***	-0.0461***	-0.0400***
	(0.00471)	(0.00531)	(0.00341)	(0.00456)
POST	0.1059***	-	-	-
	(0.00295)			
POST_REP	0.00505***	-0.0106***	-0.00972***	-0.00975***
	(0.00238)	(0.00165)	(0.00175)	(0.00172)
MALE				-0.0159***
				(0.000664)
AAPI				0.0212***
				(0.00397)
WHI				-0.0190***
				(0.00198)
BLA				0.0347***
				(0.00451)
NAT				-0.0532***
				(0.00787)
MIX				0.0293***
				(0.00247)
HISP				-0.0242***
				(0.00217)
LN_AGE				0.194***
				(0.00835)
Constant	0.545***	0.600***	0.596***	-0.174***
	(0.00657)	(0.00589)	(0.00105)	(0.0353)
Election FE	NO	YES	YES	YES
Zip Code FE	NO	NO	YES	YES
Demographic Controls	NO	NO	NO	YES
Observations	20,497,852	20,497,852	20,497,852	20,497,852
R-squared	0.0143	0.135	0.155	0.170

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (1). Treatment group: Republican voters. Control group: Democratic and Independent voters. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age)

Appendix Table A7: Pooled Difference-In-Differences, Votes-By-Mail (Including Independent Voters)

Variables	(1)	(2)	(3)	(4)
REP	0.000732***	0.000696	0.00399***	-0.00214***
	(0.000701)	(0.000715)	(0.000951)	(0.000735)
POST	0.0784***	-	-	-
	(0.00322)			
POST_REP	-0.0520***	-0.0525***	-0.0519***	-0.0517***
	(0.00187)	(0.00221)	(0.00219)	(0.00219)
WHI				-0.0190***
				(0.00221)
AAPI				0.0234***
				(0.00255)
BLA				-0.0432***
				(0.00250)
NAT				-0.0274***
				(0.00210)
MIX				-0.00547***
				(0.00188)
HISP				-0.0148***
				(0.00180)
MALE				-0.00638***
				(0.000289)
LN_AGE				0.0369***
				(0.00178)
Constant	0.0226***	0.0626***	0.0614***	-0.0600***
	(0.00107)	(0.00271)	(0.000301)	(0.00769)
Election FE	NO	YES	YES	YES
Zip Code FE	NO	NO	YES	YES
Demographic Controls	NO	NO	NO	YES
Observations	20,497,852	20,497,852	20,497,852	20,497,852
R-squared	0.004	0.075	0.085	0.089

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (1). Treatment group: Republican voters. Control group: Democratic and Independent voters. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age).

Appendix Table A8: Pooled Difference-In-Differences, Early Voting (Including Independent Voters)

Variables	(1)	(2)	(3)	(4)
REP	-0.0563***	-0.0553***	-0.0501***	-0.0378***
	(0.000400)	(0.000403)	(0.000415)	(0.000428)
POST	0.0253***	-	-	-
	(0.000238)			
POST × REP	0.0478***	0.0424***	0.0422***	0.0420***
	(0.000400)	(0.000395)	(0.000398)	(0.000397)
WHI				-0.0348***
				(0.00233)
AAPI				-0.0370***
				(0.00275)
BLA				0.0431***
				(0.00235)
NAT				-0.0605***
				(0.00339)
HISP				-0.00947***
				(0.00118)
MALE				-0.00947***
				(0.000317)
LN_AGE				0.157***
				(0.000499)
Constant	0.523***	0.537***	0.535***	-0.0788***
	(0.000229)	(0.000192)	(0.000194)	(0.00299)
Election FE	No	Yes	Yes	Yes
Zip Code FE	No	No	Yes	Yes
Demographic Controls	No	No	No	Yes
Observations	21,158,464	21,158,464	20,497,852	20,497,852
R-squared	0.003	0.072	0.092	0.104

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Data from NCSBE. DD coefficients are from the estimation of Equation (1). Treatment group: Republican voters. Control group: Democratic and Independent voters. Standard errors are clustered at the city level and are in parenthesis. Controls include interactions between election indicators and party affiliation, fixed effects by election and geographical location (Zip Code), and demographic variables (Male/Female, Asian-American Pacific Islanders, Black, Native Americans, Mixed Races, Other, Hispanic/Latino, Age).

9 Works Cited

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1), 1–19. https://doi.org/10.1111/0034-6527.00321

Atske, S. (2020, June 2). *In Changing U.S. Electorate, Race and Education Remain Stark Dividing Lines. Pew Research Center - U.S. Politics & Policy*. Retrieved December 1, 2022, from https://www.pewresearch.org/politics/2020/06/02/in changing-u-s-electorate-race-and-education-remain-stark-dividing-lines/

Atsusaka, Y., Menger, A., & Stein, R. M. (2019). Compositional Effects of Vote by Mail Elections on Voter Turnout. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3722705

Auerbach, J., & Pierson, S. (2021). Does voting by mail increase fraud? estimating the change in reported voter fraud when states switch to elections by mail. *Statistics and Public Policy*, 8(1), 18–41. https://doi.org/10.1080/2330443x.2021.1906806

Baker, M., Gruber, J., & Milligan, K. (2019). The long-run impacts of a universal child care program. American Economic Journal: Economic Policy, 11(3), 1–26. https://doi.org/10.1257/pol.20170603

Barber, M., & Holbein, J. B. (2020). The participatory and partisan impacts of mandatory vote-by-mail. *Science Advances*, 6(35). https://doi.org/10.1126/sciadv.abc7685

Baringer, A., Herron, M. C., & Smith, D. A. (2020). Voting by mail and ballot rejection: Lessons from Florida for elections in the age of the coronavirus. *Election Law Journal: Rules, Politics, and Policy*, 19(3), 289–320. https://doi.org/10.1089/elj.2020.0658

Benkler, Y., Tilton, C., Etling, B., Roberts, H., Clark, J., Faris, R., Kaiser, J., & Schmitt, C. (2020). Mail-in voter fraud: Anatomy of a disinformation campaign. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3703701

Berlinski, N., Doyle, M., Guess, A. M., Levy, G., Lyons, B., Montgomery, J. M., Nyhan, B., & Reifler, J. (2021). The effects of unsubstantiated claims of voter fraud on confidence in elections. *Journal of Experimental Political Science*, 1–16. https://doi.org/10.1017/xps.2021.18

Berinsky, A. J., Burns, N., & Traugott, M. W. (2001). Who votes by mail?: A dynamic model of the individual-level consequences of voting-by-mail systems. *Public Opinion Quarterly*, 65(2), 178–197. https://doi.org/10.1086/322196

Berry, P., Fowler, G., Waldman, M., Sanders, R., & Loving, S. (2021, May 28). *Voting laws roundup: May 2021*. Brennan Center for Justice. Retrieved November 25, 2022, from https://www.brennancenter.org/our-work/research-reports/voting-laws roundup-may-2021

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275. https://doi.org/10.1162/003355304772839588

Burden, B. C., Canon, D. T., Mayer, K. R., & Moynihan, D. P. (2014). Election Laws, Mobilization, and Turnout: The Unanticipated Consequences of Election Reform. *American Journal of Political Science*, 58(1), 95–109. http://www.jstor.org/stable/24363471

Cottrell, D., Herron, M. C., & Smith, D. A. (2021). Vote-by-mail ballot rejection and experience with mail-in voting. *American Politics Research*, 49(6), 577–590. https://doi.org/10.1177/1532673x211022626

Dellavigna, S., List, J. A., Malmendier, U., & Rao, G. (2016). Voting to tell others. *The Review of Economic Studies*, 84(1), 143–181. https://doi.org/10.1093/restud/rdw056

Domonoske, C. (2016, July 29). *U.S. Appeals Court strikes down North Carolina's voter ID law*. NPR. Retrieved February 1, 2023, from https://www.npr.org/sections/thetwo-way/2016/07/29/487935700/u-s-appeals court-strikes-down-north-carolinas-voter-id-law

Election Day Integrity Act, S.B. 326, General Assembly of North Carolina Session 2021. (North Carolina 2021) https://www.ncleg.gov/BillLookup/2021/S326#:~:text=FISCAL%20NOTE-,Filed, -Download%20Filed%20in

Elul, G., Freeder, S., & Grumbach, J. M. (2017, September 1). The effect of mandatory mail ballot elections in California. *Election Law Journal: Rules, Politics, and Policy*, 16(3), 397–415. https://doi.org/10.1089/elj.2016.0390

Fisher, S. D. (2007). Change in turnout and change in the left share of the vote. *Electoral Studies*, 26(3), 598–611. https://doi.org/10.1016/j.electstud.2006.10.006

Glynn, A. N., & Kashin, K. (2017). Front-door difference-in-differences estimators. *American Journal of Political Science*, 61(4), 989–1002. https://doi.org/10.1111/ajps.12311

Griffin, R. (2021, March 22). *Perspective | Republicans want to make it much harder to vote. that strategy could backfire*. The Washington Post. Retrieved November 25, 2022, from https://www.washingtonpost.com/outlook/2021/03/22/republican-vote restriction-turnout/

Gronke, P., Galanes-Rosenbaum, E., & Miller, P. A. (2007). Early voting and turnout. *PS: Political Science & Politics*, 40(4), 639–645. https://doi.org/10.1017/s1049096507071028

Hanmer, M. J., & Traugott, M. W. (2004). The impact of voting by mail on voter behavior. *American Politics Research*, 32(4), 375–405. https://doi.org/10.1177/1532673x04263412

Hanna, R., & Oliva, P. (2015). The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City. *Journal of Public Economics*, 122, 68–79. https://doi.org/10.1016/j.jpubeco.2014.10.004

Herron, M. C., & Smith, D. A. (2014). Race, Party, and the Consequences of Restricting Early Voting in Florida in the 2012 General Election. *Political Research Quarterly,* 67(3), 646–665. http://www.jstor.org/stable/24371899

Hill, B., & Leamer, E. E. (1980). Specification searches: AD hoc inference with nonexperimental data. *Journal of the American Statistical Association*, 75(369), 252. https://doi.org/10.2307/2287437

Kaplan, E., & Yuan, H. (2020, January). Early voting laws, voter turnout, and partisan vote composition: Evidence from Ohio. *American Economic Journal: Applied Economics*, 12(1), 32–60. https://doi.org/10.1257/app.20180192

Levitt, J. (2014, August 6). *A comprehensive investigation of voter impersonation finds 31 credible incidents out of one billion ballots cast*. The Washington Post. Retrieved November 23, 2022, from https://www.washingtonpost.com/news/wonk/wp/2014/08/06/a-comprehensive investigation-of-voter-impersonation-finds-31-credible-incidents-out-of-one-billion ballots-cast/

Lockhart, M., Hill, S. J., Merolla, J., Romero, M., & Kousser, T. (2020). America's electorate is increasingly polarized along partisan lines about voting by mail during the COVID-19 crisis. *Proceedings of the National Academy of Sciences*, 117(40), 24640–24642. https://doi.org/10.1073/pnas.2008023117

McAllister, I. (1986). Compulsory voting, turnout and party advantage in Australia. *Politics*, 21(1), 89–93. https://doi.org/10.1080/00323268608401982

Meredith, M., & Endter, Zac (2016, May 14). A*ging into Absentee Voting: Evidence from Texas [Working Paper]*. Department of Political Science, University of Pennsylvania

Mulder, B. (2020, November 13). *Fact-check: Does early voting open elections up to fraud?* Statesman. Retrieved January 2, 2023, from https://www.statesman.com/story/news/politics/elections/2020/11/13/fact-check does-early-voting-open-elections-up-to-fraud/114937128/

Nadeem, R. (2022, November 18). *Republicans and Democrats move further apart in views of voting access*. Pew Research Center - U.S. Politics & Policy. Retrieved January 18, 2023, from https://www.pewresearch.org/politics/2021/04/22/republicans-and-democrats move-further-apart-in-views-of-voting-access/.

Nagel, J. H. (1988). Voter turnout in New Zealand general elections, 1928-1988. *Political Science*, 40(2), 16–38. https://doi.org/10.1177/0032 31878804000202

NCSL. (2022, July 12). *Summary table 1: States with no-excuse absentee voting*. National Conference of State Legislatures. Retrieved April 8, 2023, from https://www.ncsl.org/elections-and-campaigns/table-1-states-with-no-excuse absentee-voting

NCSL. (2022, August 30). *Early In-Person Voting* Retrieved December 30, 2022, from https://www.ncsl.org/elections-and-campaigns/early-in-person-voting

Parks, M. (2020, August 28). *Ignoring FBI and fellow Republicans, Trump continues assault on mail-in voting*. NPR. Retrieved November 23, 2022, from https://www.npr.org/2020/08/28/906676695/ignoring-fbi-and-fellow-republicans trump-continues-assault-on-mail-in-voting.

Pennycook, G., & Rand, D. G. (2021). Examining false beliefs about voter fraud in the wake of the 2020 presidential election. *MIT Libraries*. https://doi.org/10.31234/osf.io/szdgb

Saul, S., & Epstein, R. J. (2020, April 8). *Trump is pushing a false argument on vote by-mail fraud. here are the facts*. The New York Times. Retrieved November 23, 2022, from https://www.nytimes.com/article/mail-in-voting-explained.html.

S.B. 824. General Assembly of North Carolina Session 2017. (North Carolina 2017) https://www.ncleg.net/Sessions/2017/Bills/Senate/HTML/S824v7.html

Scherer, Z. (2021, April 29). *Majority of voters used nontraditional methods to cast ballots in 2020*. Census.gov. Retrieved November 29, 2022, from https://www.census.gov/library/stories/2021/04/what-methods-did-people-use-to vote-in-2020-election.html

Shino, E., Suttmann-Lea, M., & Smith, D. A. (2021). Determinants of rejected mail ballots in Georgia's 2018 general election. *Political Research Quarterly*, 75(1), 231– 243. https://doi.org/10.1177/1065912921993537

Slusky, D. (2017). Significant placebo results in difference-in-differences analysis: The case of the aca's parental mandate. *Eastern Economic Journal*, 43(4), 580–603. https://doi.org/10.1057/eej.2015.49

Thompson, D. M., Wu, J. A., Yoder, J., & Hall, A. B. (2020). Universal vote-by-mail has no impact on partisan turnout or vote share. *Proceedings of the National Academy of Sciences*, 117(25), 14052–14056. https://doi.org/10.1073/pnas.2007249117

Viebeck, E. (2020, June 8). M*inuscule number of potentially fraudulent ballots in states with Universal Mail Voting Undercuts trump claims about election risks*. The Washington Post. Retrieved November 25, 2022, from https://www.washingtonpost.com/politics/minuscule-number-of-potentiallyfraudulent-ballots-in-states-with-universal-mail-voting-undercuts-trump-claims about-election-risks/2020/06/08/1e78aa26-a5c5-11ea-bb20- ebf0921f3bbd_story.html.

The Unintended Consequences of Ohio's Environmental Remediation Law of the Oil and Gas Industry

Abstract

This paper investigates the unintended consequences of oil and gas legislation aimed at promoting the capping of idle and orphaned wells in Ohio. On September of 2018, Ohio passed House Bill 225, which reallocated the share of oil and gas severance tax dollars, by shifting the share of funds set aside for well capping from 14% to 30%. Using a difference-in-differences comparison of the set of wells located in border counties in Pennsylvania and Ohio, this paper reflects on the effects of HB225. Findings show that while the number of well permits issued per county per month in the set of Ohio border counties increased after HB225, the set of existing wells that were plugged remained unchanged. Using a linear probability model, the probability that wells that were permitted were actually drilled remained unchanged, while the probability that drilled wells were plugged also remained unchanged overall, and actually decreased in the northern region of border counties. The findings of this study contribute to a growing literature on the failure of regulatory policy to provide meaningful effects on environmental outcomes, due to a lack of implementation strategy and unintended consequences of new legislation.

Keywords: Tax, Growth, Industry Regulation, Environmental Regulation, Orphan Well, Well Capping

JEL Codes: L71, Q35, Q52

Content

Many states in the US with significant extractable natural resources struggle (through legislation) to balance economic concerns against environmental ones. As a result, the policy decisions of states can often be imperfect, and lead to unforeseen consequences. This is the case in Ohio regarding the state's efforts to plug orphaned and idle wells.

Oil and gas wells, when left unplugged, can have serious environmental consequences. In addition, unplugged and abandoned idle wells are no longer producing and do not provide any economic gains that could offset the environmental externalities. These wells, referred to as orphans, are wells that have been abandoned and for which no owner/operator has been documented or discovered. The Interstate Oil and Gas Compact Commission (IOGCC) documents the definition of orphaned and idle wells in each state. The working definition for this paper is described in the extension Table 8 of the results section.[1] Orphan wells are severely under-reported and, if not plugged and capped, will cause methane emission, soil degradation, and leakage of toxic substances. For instance, Bussewitz and Irvine (July 31, 2021) report that unplugged orphan wells leak 5000 times more methane than plugged wells. States are able to defray some of the costs of plugging these wells through severance taxes, property taxes, and policy choices. Nelson and Fisk (2021), Bang and Hollibaugh Jr. (2021) evaluate and compare the relative efficacy of the various tools states have implemented to adequately address the costs of oil/gas capping and clean up. Legislative solutions often have to balance the interests of large industries and their economic engines against the perceived interest in environmental protection. Governing entities are wary of increasing financial assurance requirements that would insure industry responsibility for clean up because they are often seen as having a negative impact on long-term economic development. Ohio attempted to pass legislation that would provide more funds for well plugging, but would not adversely impact the economic benefits that the oil and gas industry brings to the state. However, balancing the two interests may have impacted the effectiveness of the policy decision.

This paper investigates Ohio's HB225, which earmarked a larger share of the state's severance tax imposed on oil and gas well drilling permits to be used by Ohio's oil and gas well documenting and capping program. The objective of this bill was to cap more idle and orphaned wells in the state. However, findings show that in Ohio, there was no significant impact

1 The IOGCC orphan well definition list by state is in the Federal Idle and Orphaned Oil and Gas Wells: State and Provincial Regulatory Strategies 2021 (https://iogcc.ok.gov)

on the overall number of wells plugged (those classified as working or orphaned) during the period between 2018 and 2022. At the same time, there was an increase of 3 permits issued per county-month in Ohio after 2018, which led to more drilling in the state. While the bill's objective was to increase the number of wells being plugged, it also led to the unintended consequence of a greater number of well permits being issued for new wells. This unintended consequence may have been partially a result of regulatory capture during the legislative process.

Legislative policy that aims to protect the environment can have unintended consequences that often perpetuate and even increase the problems they are designed to solve. As Stigler (1971) argued, regulatory efforts are generally captured and operated by interest groups that bend the process to conform with their own interests.[2] Peltzman (1976), Brunner, Hoen and Hyman (1983) offer a follow-up view of regulation in 1976, arguing that regulators are generally politicians who seek to maximize political capital. Interest groups might offer political support in exchange for regulation. Political actors therefore negotiate between industry focused on regulatory capture and interest group coalitions that strive for meaningful regulation. This system is highly inefficient and requires constant trade-offs. In citing Peltzman, Kroszner (2015) suggests such trade-offs "provided an empirical basis for analyzing the effectiveness of regulation and the consequences of proposed regulatory changes." Peltzman ultimately argued that trade-offs often produced unintended consequences, unforeseen costs (financial, social, political), or the creation of new problems. These then have to be factored into the overall costs of the regulatory legislation.

The Ohio legislature passed HB225 to change the Oil and Gas Well Fund allocation for abandoned well plugging from 14% to 30%.[3] Initially, the bill (as passed by the house) targeted a reappropriation of funds from

2 Stigler argues that regulation is "designed and operated primarily" for the benefit of industry. That is, regulation more often serves the interests of those being regulated than it does the interests of those whom regulation is purportedly designed to protect (e.g. consumers).

3 In addition, HB225 also provided landowners with a tax incentive by requiring the head of the ODNR "to pay the contractor for the cost of plugging and restoration, rather than requiring the Chief to reimburse the landowner after the landowner has paid the contractor as in prior law". See (September 28, 2018). HB225 waived the permit application fee for landowners for plugging their idle wells. These payment breaks provided landowners with incentives to come to the state to report and find well contractors, which greatly increased the burden on the state. It also may have caused a sudden influx of landowners looking for plugging contracts through the state as a result of the bill.

14% to 45%. The fact that oil and gas interests supported a higher appropriation may hint at regulatory capture in the state's decision-making process. Oil and gas interests represented by the Ohio Oil and Gas Association (OOGA) and Artex Oil Company argued during hearings that the ODNR had not been using their severance tax allocation exclusively for plugging orphaned wells.

Instead, they argued that the ODNR expended funds on unrelated legal settlements and plugging budget holes. Oil and gas proponent Stewart (February 28, 2018) argued for the larger allocation, with the stipulation that all those funds would be used solely for "expenses critical and necessary for protections related to oil and gas production."[4] That is, industry lobbyists accused the government of not applying proper oversight to the expenditure of well-capping funds and argued that the ODNR should be forced to spend the entire allocation exclusively on plugging wells, and not on any administrative or budget items. The bill was passed with the lower allocation but with this restrictive language in place. Given that Newell and Raimi (2018) findings show that Ohio's severance tax is one of the lowest of any state, it seems that the oil and gas industry was effectively trying to shift blame for the slow rate of plugging onto the government, and deflect any suggestions that the severance tax should be raised.[5]

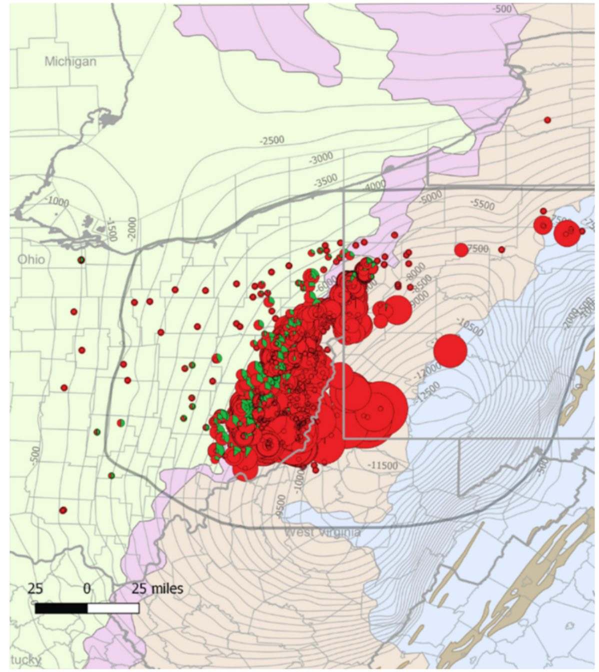
Using a difference-in-differences framework, this study engages a county-level comparison between Ohio and Pennsylvania, where Pennsylvania counties are the control group. Pennsylvania is currently the only state which does not have a severance tax in place, and therefore does not have a state controlled program for plugging wells.[6] Similar geographic topology and shale formations within the regions of interest make a difference-in-differences framework consistent. In addition, unlike Ohio,

4 See Stewart (February 28, 2018) (5), https://www.legislature.ohio.gov/legislation/132/hb225/committee

5 It almost seems as though the bill was set up to have the ODNR fail. In fact, that is exactly what ODNR resource management chief Simmers (January 16, 2018) stated in his non-proponent testimony that "if you consider the amounts appropriated in the recently passed budget, this is more than twice the amount given to the Division for operational purposes and it is six times more than what was appropriated for orphan well plugging. Simply put, there is not enough time or manpower available to accomplish this spending mandate.

6 Instead, the state provides incentives for plugging contractors through the Good Samaritan Act, which offers liability relief to well plugging contractors that have agreed to pay the costs and provide equipment for plugging a well for which they are not responsible. In addition, the Growing Greener Fund provides the Department of Energy of Pennsylvania (DEP) with funding to make grants for plugging wells.

Figure 1: Utica-Point Pleasant Shale Region Production Map



Notes: This figure shows the relative amount of oil and gas production that occurred in the Utica-Point Pleasant and Marcellus Shale regions through April, 2017. Green shaded circles represent oil production in barrels per day, while red shaded circles represent gas production in barrels of oil equivalent per day. The largest circles had greater than 2000 BPD while the smallest circles had less than 200 BPD. The map indicates that the majority of production in BPD occurred in southeastern Ohio and Western Pennsylvania. This share of production is centered around the Marcellus Shale Gas Play, which is the largest gas producing region in the northeast. Source: Energy Information Administration

Pennsylvania has not passed any legislative changes regarding the plug ging of orphaned and abandoned wells in the last decade. Therefore a county-by-county difference-in-differences comparison reveals how effective or ineffective Ohio's legislation has been. The model is similar to Eubanks and Mueller (1986) prior work at the well level in Oklahoma border counties, where Oklahoma's forced pooling law had the effect of increasing petroleum well drilling in the state.[7] This work differs from

7 Oklahoma's forced pooling increased the likelihood of a well being drilled on an acreage unit used for purchase among multiple operators. In essence, pooling allowed owners to go to the state in order to force landowners to sign contracts allowing wells to be drilled at or near their lands – known as compulsory forced pooling through incentives from the state given to these landowners – or threats. Landowners, who prior to the law, could voluntarily decide whether to allow drilling on property, were now incentivized by the states to sign contract agreements that allow operators the privilege of drilling.

Eubanks and Mueller (1986) in terms of the type of regulation studied and in that it compares border counties in neighboring states.

In measuring the effects of HB 225, this study shows that in Ohio about 3 more well drilling permits were issued per county-month after the passage of Act 225 in 2018, while no significant shift was found in the number of wells plugged. However, there was an increased probability that specifically an orphan well would be capped after the bill was passed. After 2018, it was found that older wells and shallower wells were plugged in the county-month comparison, while the probability that older and deeper wells were plugged increased. The latter result could signify that there was greater emphasis on plugging wells that were more expensive. Neither the probability of a well being plugged nor that a permitted well was actually drilled shifted significantly. Thus, while the passage of Act 225 had a significant impact on the kinds of wells that were plugged in Ohio border counties, there was an influx in permits issued following its passage. As a result, it is unclear whether the legislation had a positive net impact on environmental quality.[8]

This paper contributes to the growing literature on orphan well plugging and capping across the United States. It also offers insight into the legislative approaches that have been used to force the industry to take appropriate cleanup measures. Many studies have focused on orphan and idle well plugging and documentation, while others have examined the unintended consequences of environmental policy at the state level, and others have explored legislative mechanism failures. Newell and Raimi (2016) explain how shale counties in Ohio "have experienced a range of new revenues and costs associated with a rapid increase in shale development" which have impacted the set of interest groups that influence policies surrounding the industry. The low share of production revenues collected by the state, as described by Newell and Raimi (2018) suggests that oil and gas lobbying efforts are very prevalent and most likely influence any governing measures passed. Ohio was one of the primary states influenced by the Utica shale boom, and due to its severance tax and natural resource control and allocation, it has a measurable source of environmental cleanup run by the government.

8 After passage of the law, the Ohio Department of Natural Resources (ODNR), the agency responsible for identifying and capping abandoned and orphaned wells, was audited in August 2022 because it had failed to meet the statutory requirement for annual spending specified in the bill. Funds had accumulated in an account and remained unused. That is, the department did not cap wells at the expected rate, given the amount of funding it received, nor did it make full use of the funds available. One possible reason for this is that the legislation failed to outline a clear and meaningful plan of implementation for the Division of Oil and Gas Resources Management in the ODNR. The current Chief asserted that one reason for the lag was an inability to find available well plugging contractors with appropriate experience. Another is a delayed implementation due to the pandemic caused by Covid. See Wagoner (August 18, 2022)

1 Background

A Framework for Conducting Analysis

The problem of identifying and capping leaking abandoned and orphaned oil wells has largely become an economic burden to individual states and their tax payers. States often rely on oil and gas severance taxes, which target the extraction of natural resources intended for consumption in other states, to pay for clean up. Such taxes are supposed to internalize the expected costs for industry operations, but the funds from the severance tax often fall short of covering these costs. This was true in Ohio, for example, as Newell and Raimi (2018) found that "...a relatively low severance tax rate coupled with modest property tax collections and limited state and federal leasing led to the lowest collections of any state in FY 2013—roughly 1% of production value." This may be why the OOGA and Artex were strong proponents of HB 225, but also argued that the allocation should be higher than in the bill finally passed by the senate.

As Raimi et al. (2021) and Kang et al. (2021) show, the costs of well plugging are going up every year while revenue flatlines or declines. In the past decade research has focused on environmental costs of orphan wells and on outlining the problems facing both federal and state agencies in cleaning up and plugging abandoned wells across the US. But it has only been in the last few years that research has focused on the increased cost burden shared by states, as part of evaluating the effects of state legislation designed to confront the problem of plugging wells. Many of these studies contain valuable data compilations and analysis of existing databases. Boutot et al. (2022) contains the most up-to-date information on existing orphan well-level data across the country. Importantly, the authors suggest that newly documented orphaned wells as of April 2022 are located in areas where many other orphan wells already exist and in regions of known historical oil and gas activity (e.g., Pennsylvania, New York, and Ohio). That is, orphan wells tend to be concentrated in clusters. As such, it may be beneficial to prioritize study of regions with a high density of documented orphaned wells or regions containing legacy wells when identifying undocumented orphaned wells across the U.S. The Ohio/Pennsylvania border is one such region.

As states struggle to find solutions to these problems, they increasingly seek to pass legislation that will help allay costs and allocate funds to the growing necessity of plugging environmentally hazardous wells. In assessing which forms of legislation will result in efficient clean up of existing abandoned wells, legislators are tasked with creating policy mechanisms that will serve the needs of political interest groups while also shifting the financial burden of clean up back onto the oil and gas industry. Even in states that require financial assurance or surety bonds, the tendency of lower asset companies to default through bankruptcy increasingly makes clean-up a state rather than a private business issue. As Mallinson et al. (2022), Layton and Sprong (2022), Boomhower (2019), and Ho et al. (2018) have shown, bankruptcy protection has had a profound impact on state liability. This greatly increases the costs of cleanup and capping to states. In Ohio, the passage of House Bill 225 in 2018 was designed to help the state in its effort to cap its large number of orphaned and idle wells. The bill was passed unanimously by the state legislature and reappropriated 16% of funds from the state Oil and Gas Well Fund to be used for the purpose of identifying and plugging existing wells. Specifically, the ODNR was required to increase the share of funds allocated from the severance tax towards plugging abandoned wells from 14% to 30%. However, there was an audit on the ODNR in August of 2022 to determine why the department failed to use all of its funds available for plugging.

The bill's relative effectiveness will be determined by whether the number of permits issued for drilling per month stayed the same, decreased, or increased after 2018. The passage of the bill could potentially lead to moral hazard if companies are incentivized to drill more new wells in Ohio knowing that orphan wells are more likely to be plugged by the state due to the reallocation of funds. As explained by London-based, non-profit think tank, Carbon Tracker's 2020 orphan well report Schuwerk and Rodgers (2020), states have inadvertently incentivized oil and gas companies to hold onto their wells for as long as possible before permanently abandoning them. In addition, companies with more assets will often sell their oldest wells to companies with weaker marginal assets, which makes it more likely that wells will be abandoned down the road through bond forfeiture and bankruptcy. This means that the number of self-bonded and abandoned wells is exponentially expanding rather than decreasing. Ohio's HB225 was passed to address these issues of financial responsibility for plugging the increasing number of abandoned wells. This paper is designed to assess the success of the bill in addressing this issue. It does this by measuring the possible correlation between higher

levels of funding and an increased share of wells plugged. As a corollary, the paper also asks if the bill's passage had an effect on the number of permits issued after 2018, as this may be endogenous with the passage of the bill because of the reasons described above.

B Existing Data

Most existing state databases on oil and gas wells are incomplete due to the policy aims of state governments and the lack of federal funding targeting industry cleanup. The main data sources for this paper come from Pennsylvania's Department of Environmental Protection (DEP) and the ODNR's oil and gas well databases which contain information on plugging. In order to merge the two state datasets with reliable existing well-level data, the study takes data from ShaleXP and matches data at the well level with the two state datasets. In determining how wells in both states should be categorized, I used the states' definition of orphan wells provided by the Interstate Oil and Gas Compact Commission (IOGCC). Ohio defines orphan wells as those "for which a bond has been forfeited or an abandoned well for which no money is available to plug the well." This then is the paper's working definition for comparing wells in Pennsylvania, which are identified by their historic/unknown owner and forfeit history. Both data sources include information on well depth, age, geographic location, and API number for matching.[9] However, some wells have missing entries in multiple categories, and cost estimate data for well plugging are not included in permit documentation.[10]

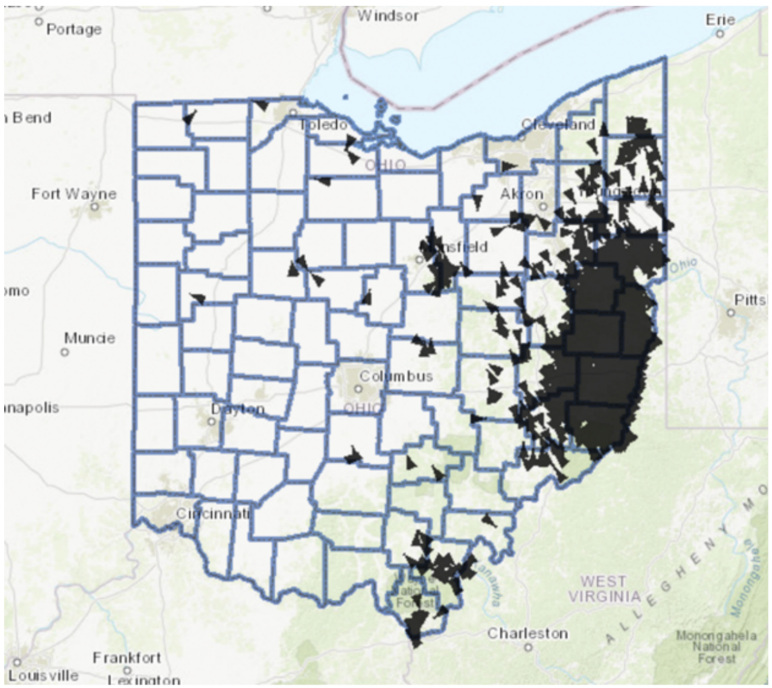
9 The purpose of the API Number is to uniquely and permanently identify every well and wellbore pertinent to the petroleum industry of the USA. Though defined for the petroleum industry, this standard may be applied to other classes of wells (e.g., water, sulfur, coal) as deemed appropriate by the regulatory agency, state, or other entity. If applied to other such classes, the uniqueness of each API Number must be maintained across the class boundaries. See (2013): API Number Standard: An Identifier for Petroleum Industry Wells in the USA (6)

10 One existing source of data not used in the analysis was Enverus DrillingInfo rig-level data provided through PRISM software. This data includes comprehensive well level data across the country to the oil and gas industry, and has been used by other literature: Kang et al. (2021), Schuwerk and Rodgers (2020). However, unfortunately access to data from Enverus required financial resources that were not available. A follow-up study using datasets provided by Enverus and comparing them to the data set used for the analysis in this paper would be beneficial and necessary for future consideration of the results section.

2 Data for Analysis

The sets of well-level data used for regression came from the ODNR's oil and gas well database, the DEP's Plugged and Abandoned lists, and well-level spreadsheet files stored in ShaleXP. The data was merged by API number. The API number is a well-identification number in the U.S. that uniquely defines any wells that were ever permitted to be drilled. Permits issued by states to drilling companies do not necessarily result in the well ever being drilled. Many wells do not have a SPUD (spudding) date since they never passed the conception stage, but in states such as Ohio where companies are required by law to obtain a permit for drilling, funds towards the state's severance tax are collected from drillers who acquire a permit before the drilling process begins. As a result, one of the larger and more secure sources of funding for Ohio's severance tax comes from the permitting process, which is why this paper focuses on the number of new well permits being issued.

Figure 2: Oil and Gas Well Drilling in Ohio



Notes: This figure shows the documented oil and gas wells drilled in the state of Ohio. The majority of drilling in the state occurs in the southeastern region, where the Utica and Marcellus shale regions are located. Source: Ohio Department of Natural Resources Interactive Oil and Gas Well Map Locator.

One of the issues across state databases and private sources is incomplete information regarding well depth, elevation, and production. The data currently used for the preliminary regressions in this paper do not include missing data that could be provided from Enverus.[11] Tables 1 and 2 provide summary statistics of the combined and merged well-level dataset used for this analysis. The data in table 1 includes 4,612 unique permits issued from 2014-2022, with 5,163 county-month entries (of these, 551 county-month entries had 0 well permits issued, as matched in this dataset). While some wells were both permitted and plugged from 2014-2022, the majority of wells that were plugged were drilled prior to 2014. Well-level data in table 2 includes some of the data in table 1, but matches data that includes the set of wells that were drilled at any point from 1800 to the present (most wells from Pennsylvania were drilled much earlier than those in Ohio, but Pennsylvania records regarding earlier drilling prior to 1950 are estimated and often show up in the records as nothing more specific than 1800). The wells for which no ownership or existing permit data is specified are considered idled or orphaned as by the definition used in this paper and discussed earlier. Of the wells included in table 2, 2,235 wells were plugged from 2014-2022 (782 Ohio wells and 1,453 Pennsylvania wells).

Table 1: Summary Statistics: Permit Dataset

Variable Name	of Observations	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)	(5)
Average Oil & Gas Price	n = 4,612	29.37	8.121	11.10	61.69
Current Oil Production (Bpd)	n = 4612	7	149	0	5215
Current Gas Production (Btu)	n = 4612	96,863	239,037	0	23,900,000
Drillers Total Depth (ft)	n = 4,612	3,708	7,734	0	29,192
Ground Level Elevation (ft)	n = 4,612	838	605	0	1584
Permit Count Per County-Month	n = 1,152	18.34	15.292	0	65
Ohio Well Permits Issued	n = 1,376				
Pennsylvania Well Permits Issued	n = 3,236				
Total Well Permits Issued in Data Set	n = 4,612				

Notes: This table reports summary statistics from the set of wells permitted to be drilled ranging from September 28, 2014 to September 28, 2022.

11 Enverus DrillingInfo and Rig data that has been acquired and used by other authors: Kang et al. (2021), Schuwerk and Rodgers (2020), and Boutot et al. (2022). The Enverus database is the most comprehensive oil and gas well-level database currently available in the United States, but there are discrepancies between it and state databases as documented by Schuwerk and Rodgers in 2020.

Table 2: Summary Statistics: Plug Dataset

Variable Name	of Observations	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)	(5)
Average Oil & Gas Price	n = 2,235	31.73	10.60	8.70	61.69
Well Age (days)	n = 2,235	4,137	5,159	0	44,465
Drillers Total Depth (ft)	n = 2,235	508	1,603	0	24,146
Ground Level Elevation (ft)	n = 2,235	434	598	0	2500
Plug Count Per County-Month	n = 1,152	3.717	3.773	0	17
Ohio Wells Plugged	n = 782				
Pennsylvania Wells Plugged	n = 1,453				
Total Wells Drilled in Data Set	n = 4,108				

Notes: This table reports summary statistics from the set of drilled wells that were plugged during the period between September 28, 2014 to September 28, 2022. 2,235 wells were plugged during the period.

3 Empirical Strategy

The regressions in this paper can be used and analyzed in three ways. The first method measures the monthly number of permits issued for new drilling and the number of wells plugged as the dependent variables with a difference-in-differences framework. Then, the probability that wells were plugged is compared to the probability that a permitted well is drilled. There are several factors that affect well plugging and drilling costs. In particular, wells are often drilled in shale basins where resource extraction is widely available, leading to clustering and an influx of drilling and plugging permits issued in specified regions. In determining whether a well is drilled or plugged, it is important to focus on factors at the well level, including depth, age, and location. In addition, controls for production are necessary when the likelihood that a well is plugged is correlated with other well fixed effects, particularly age and depth. In addition to well fixed effects, variables that affect the base of drilled and plugged wells include the prices for extractable resources and the political and economic regimes that are in place. This paper uses oil and gas

prices as well as time factors that affect the level of well capping that takes place in Pennsylvania and Ohio. The time factors which control for the major political and economic changes that occurred during the sample period is Biden's presidency after the passage of the Inflation Reduction Act, and the pandemic in 2020, where firms, particularly in Ohio, were widely shut down. The main equations for this model are as follows:

$$AvgWellsPlugged_{cm} = \beta_0 + \beta_1 OHCounty + \beta_2 Post + \beta_3 OHCountyPost + \theta X_{cm} + \sum_{n=1}^{2} \omega_{cmn} OHCountyPost + \varepsilon_{cm} \quad (1)$$

The subscripts c and m represent the averages for each term on the group of wells plugged in each county-month from 2014 to 2022. β_3 is the coefficient of interest, θX_{cm} is the set of controls in county c during month m, and ω_{cmn} is the age or depth interaction coefficient with the post period in Ohio. The set of controls includes the oil and gas price during the county-month, binary controls for whether the wells in the county-month are located in the Marcellus Shale Gas Play, whether the wells were plugged during Biden's term in office, and whether the county month was during the start of the pandemic. Price controls are important as Brown, Maniloff and Manning (2020) notes that responses to changes in taxes at the state level have differential effects to changes in national output prices for natural resources. While the state severance tax value did not shift after the passage of the law, price changes are recorded every month and need to be controlled for at the state level. Similarly, below is the regression where the number of well drilling permits issued is the dependent variable:

$$AvgPermitsIssued_{cm} = \beta_0 + \beta_1 OHCounty + \beta_2 Post + \beta_3 OHCountyPost + \theta X_{cm} + \varepsilon_{cm} \quad (2)$$

The coefficients are the same except for the omission of the interaction terms on well age and depth. In the average permit regression, the well depth is the proposed total depth as issued on the permit rather than the average depth of existing wells, as in equation (1). Most of the average proposed total depth of these wells in the given county-month are non-zero since these wells were initially planned to be drilled. Permit age is the average time from which the wells were permitted to when they were drilled, or if the wells were never drilled, then this age coefficient is the time from when the permit was issued to the time that the permit expired or was cancelled. This differs from the well age as calculated in equation (1), which represents the average age of existing wells in the given county-month. Proposed total depth and permit age are measured in the appendix as dependent variables with the same set of independent variables and controls as in (2). County and month fixed effects outside of the set of controls are not included in either regression, and interpretations of the results of this paper should take this into consideration when considering potential bias of the coefficient terms.

The same framework is conducted using linear probability models. The dependent variable is instead the likelihood that a well that was initially permitted was drilled. Similarly, it makes sense to run this as a difference-in-differences model, to determine the effect of HB225 on the probability that wells and permits are plugged and issued, respectively. The LPM model is:

$$ProbabilityPermittedWellIsDrilled_{icm} = \beta_0 + \beta_1 OhioCounty + \beta_2 PostPeriod + \beta_3 OHCountyPostPer + \theta X_{cm} + \delta_i + \varepsilon_{cm} \quad (3)$$

Rather than taking the average data for the group of wells plugged during each county month, this LPM model assesses the probability that a single well that was permitted was drilled. The coefficient on the Ohio county post period is of interest, and the subscripts i, c, and m represent well i in county c during month m. Well fixed effects are represented by δi and include the age of the well, the well type, and the proposed total depth as issued on the permit for drilling. Additional controls are the same as in equations (1) and (2).

Using the same design as the regression from above, the LPM model for the probability that a well that was drilled was actually plugged is as follows:

$$ProbabilityDrilledWellIsPlugged_{icm} = \beta_0 + \beta_1 OhioCounty + \beta_2 PostPeriod + \beta_3 OHCountyPostPer + \theta X_{cm} + \delta_i + \varepsilon_{cm} \quad (4)$$

The well fixed effects from equation (3) are the same as in (4), but the sample size is adjusted to include only wells that were drilled and not the entire base of permits issued from 2014-2022. However, this also means that some wells that were drilled prior to 2014 are included in the dataset used to run this model. As an example, a large sample of Pennsylvania wells drilled in 1800 were plugged sometime during the event study interval, and therefore are included in the regression specifications from (4).

The third set of regressions are logistic and probit. Similar to the LPM models, these models would work well with plugging probability models or permitting models that may be more accurate than a basic LPM model. These models ensure that there are no negative probabilities in measuring the dependent variable, but lack the robust set of controls that are provided in the LPM and County-Month regressions.

4 Results

A Main Result

I find evidence that more well permits were issued as a result of HB 225. Using a difference-in-differences framework, columns 1 and 2 of Table 3 report results showing that about three more permits per county were issued following the law's passage. Columns 3 through 6 suggest this is due primarily to the large increase in permits issued in southern counties near the Marcellus shale. Results are statistically significant at the 1% level.

Table 3: Well Permits Issued Per County-Month: September 2014-2022

	All Border Counties		Northern Region		Southern Region	
	(1)	(2)	(3)	(4)	(5)	(6)
OH County Post	2.941***	2.940***	0.214	0.212	5.653***	5.653***
	(0.749)	(0.753)	(0.192)	(0.192)	(1.474)	(1.482)
Ohio County	-3.734***	-3.735***	0.342**	0.342**	-9.597***	-9.597***
	(0.674)	(0.675)	(0.159)	(0.160)	(1.365)	(1.369)
Post Period	-4.733***	-3.445***	-0.521	-0.269*	-8.944***	-6.631***
	(0.697)	(0.745)	(0.133)	(0.160)	(1.373)	(1.475)
Near Marcellus Shale Gas Play	Yes	Yes	Yes	Yes	No	No
during Biden Control	No	Yes	No	Yes	No	Yes
Oil And Gas Price Controls	No	Yes	No	Yes	No	Yes
Pandemic (2020) Control	No	Yes	No	Yes	No	Yes
n =	1152	1152	576	576	576	576
R^2	0.208	0.214	0.065	0.077	0.217	0.230

Notes: This table reports the regressions of permits issued per county per month on the difference-in-differences treatment variables and a set of controls. Even number columns include time and price controls. Columns (3) and (4) restrict the sample to county months observed only in the northern region of border counties, while columns (5) and (6) restrict the sample to only the county months observed in the southern region of border counties. Standard errors clustered at the month level. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

Table 4 shows the number of wells plugged per county-month for all border counties and the separate regions. Even number columns include the full set of price, production, and time controls. The effect of HB225 on the number of wells plugged in Ohio counties is negligible in all specifications. During the post period in Ohio, when the average depth of wells was lower per county per month, the number of wells plugged was higher. This finding is represented by row 2 with the Well Depth/Ohio County Post Period interaction effect. The number of wells plugged per county per month was more overall when the average well age in Ohio counties during the given month was greater (less in the northern region but more in the southern region). This is captured by the interaction term in row 3. The border counties in both states showed significant positive effects on the number of wells plugged depending on the average depth of wells plugged and the age of wells plugged per county per month at the 1 percent level.

Table 4: Wells Plugged Per County-Month: September 2014-2022

	AllBorderCounties		Northern Region		Southern Region	
	(1)	(2)	(3)	(4)	(5)	(6)
OH County Post	0.016	0.057	0.248	0.248	-0.462	-0.395
	(0.251)	(0.192)	(0.174)	(0.171)	(0.280)	(0.277)
Well Depth * OHCountyPost	-0.283***	-0.351***	0.048	0.027	-0.285**	-3.752***
	(0.078)	(0.069)	(0.074)	(0.071)	(0.115)	(0.125)
Well Age * OHCountyPost	0.015***	0.018***	-0.021**	-0.018**	0.019***	0.021***
	(0.0046)	(0.0041)	(0.0079)	(0.0081)	(0.0045)	(0.0054)
Average Well Depth (Thousands of ft)	0.361***	0.351***	0.078	0.074	0.346***	0.343***
	(0.064)	(0.051)	(0.050)	(0.051)	(0.112)	(0.114)
Average Well Age (yrs)	0.014***	0.014***	0.048***	0.045***	0.0087***	0.0090***
	(0.0013)	(0.0017)	(0.0058)	(0.0059)	(0.0018)	(0.0019)
Ohio County	-1.402***	-1.361***	-0.310**	-0.288**	-2.265***	-2.210***
	(0.179)	(0.133)	(0.119)	(0.117)	(0.183)	(0.186)
Post Period	0.128	0.196	-0.019	-0.019	0.557**	0.472
	(0.164)	(0.225)	(0.146)	(0.185)	(0.277)	(0.349)
Marcellus Gas Play Control	Yes	Yes	No	No	Yes	Yes
Oil and Gas Price Controls	No	Yes	No	Yes	No	Yes
During Biden Control	No	Yes	No	Yes	No	Yes
Combined Oil and Gas Production Control	No	Yes	No	Yes	No	Yes
Pandemic (2020) Control	No	Yes	No	Yes	No	Yes
n =	1152	1152	576	576	576	576
R^2	0.279	0.292	0.371	0.380	0.280	0.288

Notes: This table reports the regressions for the number of wells plugged per county per month on the difference-in-differences treatment variables, well age, depth, and their interactions with the post period, and a set of controls. Even number columns include time, price, and production controls. Columns (3) and (4) restrict the sample to county months observed only in the northern region of border counties, while columns (5) and (6) restrict the sample to only the county months observed in the southern region of border counties. Standard errors clustered at the month level Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

Tables 5 and 6 show the results of the linear probability model regressions on the difference-in-differences treatment variables and a set of well and time controls. The probability that a permitted well saw actual drilling occur did not shift significantly after the passage of HB225 in Ohio, except in the northern region where there was nearly a 50% increase in the probability in column (4). The data in (3), (5), and (7) excludes stratigraphic test wells. Stratigraphic tests wells are wells that are drilled solely for the purpose of exploring possible extraction in a region, without the intent of actually producing any oil or gas from the well itself. In particular, the top entry of column (4) suggests that there was a large influx in test wells drilled (nearly a 50% increase) in the Northern region in Ohio after the passage of HB225. In addition, a one-foot increase in the proposed depth of the well significantly increased the probability that the permitted well was drilled in all specifications. This makes sense, as large well projects that have deeper drilling are more expensive and can only be taken on by larger companies that are more financially stable than smaller companies, which are more likely to drill smaller wells. In addition, in all specifications, a well that was designed to drill for both oil and gas saw a much smaller probability of actually being drilled compared to the base group, which was the set of gas-only producing wells. Intuitively, this could mean that combined oil and gas wells are much more difficult projects, leading to more cancellations and expiration in well drilling permits for these kinds of wells.

Table 5: Probability that Permitted Well is Drilled (LPM Model)

	AllBorderCounties			Northern Region		Southern Region	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
OH County Post	0.013	0.015	0.001	0.485**	0.042	0.043	0.027
	(0.055)	(0.055)	(0.054)	(0.213)	(0.116)	(0.056)	(0.056)
Proposed Total Depth (Thousands of ft)	0.023***	0.024***	0.024***	0.103***	0.104***	0.024***	0.024***
	(0.0025)	(0.0025)	(0.0026)	(0.020)	(0.020)	(0.0026)	(0.0027)
Combined Oil and Gas Well	-0.457***	-0.437***	-0.166**	-0.637***	0.085	-0.414***	-0.002
	(0.082)	(0.082)	(0.082)	(0.236)	(0.098)	(0.082)	(0.092)
Alternative Well Type	0.083	0.075	-0.007	0.377**	-0.090	0.069	0.050
	(0.053)	(0.052)	(0.044)	(0.187)	(0.097)	(0.051)	(0.048)
Ohio County	-0.560***	-0.547***	0.046	-0.832***	-0.344	-0.542***	-0.128
	(0.063)	(0.061)	(0.090)	(0.168)	(0.230)	(0.059)	(0.103)
Post Period	0.068	0.056	0.061	-0.484**	-0.036	0.041	0.048
	(0.044)	(0.053)	(0.053)	(0.217)	(0.104)	(0.054)	(0.054)
Marcellus Gas Play Control	Yes	Yes	Yes	Yes	Yes	No	No
During Biden Control	No	Yes	Yes	No	Yes	No	Yes
Oil And Gas Price Control	No	Yes	Yes	No	Yes	No	Yes
Pandemic (2020) Control	No	Yes	Yes	Yes	Yes	Yes	Yes
n	4,612	4,612	4,464	242	242	4,370	4,222
R-Square Adj	0.161	0.161	0.119	0.665	0.650	0.154	0.107

Notes: This table reports the linear probability model regressions where the binary dependent variable is the probability that a well which received a permit was actually drilled. Column 1 includes the full set of permits issued from 2014-2022, while column (2) adds time and price controls. Columns (4) and (6) match column (2) but restrict the sample to the given regions. Columns (3) restricts the sample to exclude test wells (wells drilled for exploration and not for production) drilled during the period. Columns (5) and (7) match column (3) but restrict the sample further to the region level. Standard errors clustered at the month level. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

The models for the probability that permitted and/or actually drilled wells were plugged are reported in table 6. Specifications (1), (4), and (6) include the set of wells that were permitted between 2014 and 2022 and exclude wells that were drilled or permitted prior to 2014. Specifications (2), (3), (5), and (7) include the base of all wells that were drilled at any point, and excludes any permits issued for wells which were not drilled. This set includes wells that were drilled prior to 2014 and which are considered idle or orphan if there is no existing permit documentation tied to the drilled well. All specifications with the exception of (5) saw a negative shift in the probability that a well was plugged, particularly in the southern region where the probability that a well was plugged in Ohio counties after 2018 shifted 17% and 36% downward in (6) and (7), respectively. Out of the set of wells which were drilled, there was actually a downward shift overall in the probability that a well was plugged in all Ohio border counties after HB225 (shown in the top entry of column (3)). The interaction effects on well age and depth in all entries except for the northern region depth entries show positive effects, implying that deeper and older wells were more likely to be plugged after HB225. Although the overall downward shift in plugging probability suggests unintended consequences of the bill, it does suggest that funds were used to plug wells that are more expensive to plug, particularly those that are old and deep.

Table 6: Probability that Permitted Well is Plugged (LPM Model)

	AllBorderCounties			Northern Region		Southern Region	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
OH County Post	-0.170***	-0.131	-0.128*	-0.528	0.0253	-0.165**	-0.363***
	(0.062)	(0.079)	(0.077)	(0.356)	(0.160)	(0.068)	(0.088)
Well Age * OH County Post	0.0094***	0.0030***	0.0030***	-0.0070	-0.0019**	0.0091***	0.0059***
	(0.0022)	(0.00078)	(0.00075)	(0.0087)	(0.00089)	(0.0024)	(0.0017)
Well Depth * OH County Post	0.0087***	0.0089**	0.0096***	0.095	0.017	0.0086***	0.021***
	(0.0030)	(0.0036)	(0.0037)	(0.063)	(0.033)	(0.0033)	(0.0042)
Combined Oil and Gas Well	-0.706***	-0.591***	-0.570***	-0.598**	-0.170**	-0.697***	-0.616***
	(0.081)	(0.020)	(0.019)	(0.288)	(0.081)	(0.081)	(0.020)
Alternative Well Type	0.438***	0.163***	0.163***	0.376*	-0.0621	0.442***	0.165***
	(0.052)	(0.023)	(0.023)	(0.215)	(0.109)	(0.051)	(0.024)
Well Age	0.0030***	0.0023***	0.0023***	0.0052**	0.0017**	0.0029***	0.0023***
	(0.00025)	(0.00018)	(0.00017)	(0.0021)	(0.00063)	(0.00025)	(0.00017)
Drillers Total Depth (Thousands of ft)	-0.016***	-0.021***	-0.0022***	-0.0038	-0.049	-0.015***	-0.030***
	(0.0026)	(0.0027)	(0.00028)	(0.027)	(0.028)	(0.0028)	(0.0031)
Ohio County	-0.414***	0.335***	0.335***	-0.241	0.010	-0.420***	0.525***
	(0.053)	(0.057)	(0.057)	(0.261)	(0.170)	(0.053)	(0.063)
Post Period	-0.009	-0.016	-0.044**	0.538**	0.038	-0.015	-0.057***
	(0.015)	(0.016)	(0.019)	(0.228)	(0.069)	(0.016)	(0.020)
Oil and Gas Price Controls	Yes	No	Yes	Yes	Yes	Yes	Yes
Combined Oil/Gas Production Control	No	No	Yes	No	Yes	No	Yes
Marcellus Gas Play Control	Yes	Yes	Yes	Yes	Yes	No	No
During Biden Control	Yes	No	Yes	Yes	Yes	Yes	Yes
Pandemic 2020 Control	Yes	No	Yes	Yes	Yes	Yes	Yes
n =	4,612	4,108	4,108	242	494	4,370	3,614
R-Square Adj	0.481	0.655	0.667	0.382	0.276	0.484	0.690

Notes: This table reports the linear probability model regressions where the binary dependent variable is the probability that a permitted well is plugged. Column (1) includes the full set of permitted wells from 2014-2022. Columns (3) and (5) are identical to column (1) except for the restriction by region. Column (2) restricts the sample of permitted wells to only include drilled wells, and column (3) does the same but includes the set of time, production, and price controls. Columns (5) and (7) are identical to column (3) but restrict the sample by region. Standard errors clustered at the month level. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

B Extensions and Robustness

Table 7 reports the logistic and probit models for comparison with the LPM models in tables 5 and 6. The reported coefficients of interest are the three interactions, measuring the effect on the probability that a well is plugged. The odd number specifications are logit regressions, while the even number specifications are probit models. Average marginal effects of the interaction terms are reported. There is no significant shift in the probability that a well is plugged in the southern region in Ohio after 2018, but a significant downward shift in the northern region. There are negligible differences in the probability margins of the interaction between well depth and OH County Post, and a slight positive marginal difference for the interaction between age and OH County Post (about 0.5% higher for specifications 1 and 2 that include all border counties), but not in the northern region.

Table 7: Probit and Logit Models for The Probability of Wells Plugged

	Logistic	Probit	Northern Region		Southern Region	
	(1)	(2)	(3)	(4)	(5)	(6)
OH County Post (Marginal Difference)	-0.0088	0.00069	-1.937***	-1.388***	-0.055	-0.063*
	(0.035)	(0.039)	(0.434)	(0.366)	(0.039)	(0.037)
Well Depth * OH County Post (Marginal Difference)	-0.000082	-0.0039	-0.050	-0.046	-0.0015	-0.0011
	(0.0053)	(0.0042)	(0.068)	(0.054)	(0.0027)	(0.0027)
Well Age * OH County Post (Marginal Difference)	0.0043***	0.0043***	-0.0031	-0.0030	0.0078*	0.0073***
	(0.0011)	(0.0012)	(0.0038)	(0.0031)	(0.0040)	(0.0019)
Oil and Gas Price Controls	Yes	Yes	Yes	Yes	Yes	Yes
Plugging Cost Control	Yes	Yes	Yes	Yes	Yes	Yes
Combined Oil and Gas Production Control	Yes	Yes	Yes	Yes	Yes	Yes
During Biden Control	Yes	Yes	Yes	Yes	Yes	Yes
Near Marcellus Shale Gas Play Control	Yes	Yes	No	No	No	No
Pandemic 2020 Control	Yes	Yes	Yes	Yes	Yes	Yes
n =	3,382	3,382	362	362	3,012	3,012
R^2	0.619	0.608	0.327	0.327	0.655	0.642

Notes: This table reports the probit and logit average marginal effects for three interaction effects. Odd number columns are logit regressions while even numbered columns are probit regressions. Columns (3) and (4) restrict the sample to the northern region while columns (5) and (6) restrict the sample to the southern region. Average marginal effects for probit and logit regressions can be interpreted as the percent change in probability that a permitted well is plugged on average, holding other variables in the model constant. Clustered at the month level. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

As an extension to Table 6 there are roughly 1500 wells that were plugged in the regressions with the set of drilled wells (excluding stratigraphic test wells). Using this base of wells, the additional LPM model in Table 8 employs the probability that a well is orphaned as a dependent variable. The definition of orphan wells from the IOGCC differs for Pennsylvania and Ohio. The working definition used in this paper is that a well is considered orphaned if its owner is unknown, if it has a missing permit date, and/or if it is older than 100 years. Table 8 shows that during the post-2018 period in Ohio, the probability that a plugged well is orphaned is much higher than it was before 2018.[12] Deeper wells were less likely to be orphaned than shallower wells and older wells were more likely to be orphaned in the Southern region and less likely to be orphaned in the Northern region. In addition, combined oil and gas wells or alternative well types (dry holes, test holes, water wells, etc.) were more likely to be orphaned than wells that only produced gas.

Table 8: Probability that Plugged Well is an Orphaned Well

	All Border Counties	Northern Region	Southern Region
	(1)	(2)	(3)
OH County Post	0.742***	0.553*	0.594***
	(0.184)	(0.291)	(0.201)
Well Depth * OH County Post	-0.112***	-0.084*	-0.069***
	(0.022)	(0.047)	(0.018)
Well Age * OH County Post	-0.00083	-0.0033**	0.0021
	(0.0015)	(0.0014)	(0.0019)
Combined Oil Gas Well	0.111***	-0.019	0.138***
	(0.022)	(0.030)	(0.027)
Alternative Well Type	0.029	0.137	0.025
	(0.018)	(0.086)	(0.018)
Well Depth (Thousands of ft)	0.0039	-0.055	0.0062
	(0.0053)	(0.040)	(0.0047)
Well Age	0.0046***	0.0049***	0.0045***
	(0.000095)	(0.00089)	(0.00010)
OH County	-0.078	0.361	-0.207***
	(0.077)	(0.260)	(0.059)
Post Period	-0.157***	-0.113**	-0.142***
	(0.030)	(0.046)	(0.034)
Oil and Gas Price Controls	Yes	Yes	Yes
Near Marcellus Shale Gas Play Control	Yes	No	Yes
During Biden Control	Yes	Yes	Yes
Pandemic 2020 Control	Yes	Yes	Yes
n	1,461	402	1,059
R^2	0.729	0.600	0.746

Notes: This table is an extension of table 6 and reports the difference-in-differences treatment variables on the probability that a plugged well is orphaned. Columns (2) and (3) restrict the sample to only include the northern and southern regions, respectively. Clustered at the month level. Orphan Wells match the definition given for Pennsylvania in the IOGCC. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

12 In fact, this may correlate to the increase plugging of orphaned wells after the passage of HB225. According to Hicks (2019-2020), Ohio after passage of the bill ”plugged 233 wells over a twenty-two month period using sixteen million dollars from the fund.” (23). Here, he is citing in footnote 188 Graves (July 23, 2019).

Of the 1500 plugged wells, 678 wells were orphan wells with unknown owners and missing permit dates. Using this set of wells, Table 9 reports the probability that these orphaned wells were plugged following the passage of HB225. Results indicate no significant effect of the bill on the plugging of orphaned wells, which differs from the results indicated in Table 8. This may be due to the fact more orphaned wells were plugged overall in the post period relative to the pre period, but because of the greater amount of orphaned wells being discovered by the ODNR during the post period, the probability that an orphaned well was plugged may have fallen because the proportion of plugged orphaned wells to newly discovered orphaned wells decreased. Table 10 reports the county-month regressions for the number of wells plugged using the same set of only orphaned wells. Similarly to table 9, there appears to be a positive effect of the bill on the number of orphaned wells plugged (although statistical significance varies by specification, and in the southern region, there is no statistically significant effect found). Note that the results of Tables 9 and 10 do not indicate a failure of the ODNR to address orphaned wells specifically. If more orphaned wells are being documented within the state, then the program succeeded in identifying a larger problem than what was initially estimated to be. Had these newly discovered orphaned wells not been found, then the unobserved environmental costs of unplugged orphaned and idle oil and gas wells would be greater. Thus, the ODNR and the passage of HB225 helped identify a more accurate representation of the orphaned well crisis within the state.

Table 9: Probability that Orphaned Well is Plugged

	All Border Counties	Northern Region	Southern Region
	(1)	(2)	(3)
OH County Post	0.271	-0.692	-0.102
	(0.306)	(0.566)	(0.402)
Well Depth * OH County Post	-0.037	0.030	0.069
	(0.075)	(0.036)	(0.084)
Well Age * OH County Post	0.000016	0.00676	0.00374*
	(0.00085)	(0.0054)	(0.0019)
Combined Oil Gas Well	-0.232***	-0.282	-0.238***
	(0.032)	(0.242)	(0.032)
Alternative Well Type	0.119***	-0.261	0.127***
	(0.044)	(0.221)	(0.045)
Well Depth (Thousands of ft)	0.072	-0.030	0.072
	(0.0073)	(0.036)	(0.062)
Well Age	0.0010**	-0.0068	0.0011**
	(0.00042)	(0.0054)	(0.00043)
OH County	0.030	-0.033	0.028
	(0.311)	(0.157)	(0.347)
Post Period	-0.076	0.065	-0.078
	(0.065)	(0.064)	(0.064)
Oil and Gas Price Controls	Yes	Yes	Yes
Near Marcellus Shale Gas Play Control	Yes	No	Yes
During Biden Control	Yes	Yes	Yes
Pandemic 2020 Control	Yes	Yes	Yes
n	678	71	607
R^2	0.208	0.318	0.201

Notes: This table is an extension of table 6 and reports the difference-in-differences treatment variables on the probability that an orphaned well is plugged. Columns (2) and (3) restrict the sample to only include the northern and southern regions, respectively. Clustered at the month level. The set of wells measured are orphaned wells with an unknown owner and missing permit date. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

Table 10: Probability that Orphaned Well is Plugged

	AllBorderCounties		Northern Region		Southern Region	
	(1)	(2)	(3)	(4)	(5)	(6)
OH County Post	0.279	0.332**	0.137*	0.148*	0.353	0.406
	(0.177)	(0.155)	(0.074)	(0.076)	(0.337)	(0.332)
Well Depth * OHCountyPost	-0.161	-0.164	-0.833	-0.841	0.665***	0.775***
	(0.463)	(0.462)	(0.700)	(0.705)	(0.201)	(0.234)
Well Age * OHCountyPost	0.0046	0.0044	0.021	0.021	0.0028	0.001
	(0.0036)	(0.0036)	(0.020)	(0.021)	(0.0038)	(0.0041)
Average Well Depth (Thousands of ft)	0.567	0.568	0.949	0.954	0.043***	0.033*
	(0.434)	(0.431)	(0.669)	(0.674)	(0.016)	(0.018)
Average Well Age (yrs)	0.017***	0.017***	0.0046	0.0044	0.016***	0.016***
	(0.0016)	(0.0017)	(0.020)	(0.020)	(0.0021)	(0.0018)
Ohio County	-0.454***	-0.457***	0.024	0.023	-1.049***	-1.052***
	(0.112)	(0.108)	(0.024)	(0.025)	(0.218)	(0.218)
Post Period	-0.092	0.196	0.001	-0.023	-0.143	-0.497
	(0.159)	(0.145)	(0.012)	(0.054)	(0.329)	(0.306)
Marcellus Gas Play Control	Yes	Yes	No	No	Yes	Yes
Oil and Gas Price Controls	No	Yes	No	Yes	No	Yes
During Biden Control	No	Yes	No	Yes	No	Yes
Combined Oil and Gas Production Control	No	Yes	No	Yes	No	Yes
Pandemic (2020) Control	No	Yes	No	Yes	No	Yes
n =	1152	1152	576	576	576	576
R^2	0.455	0.504	0.371	0.441	0.438	0.458

Notes: This table is an extension of table 4 and reports the difference-in-differences treatment variables on the number of orphaned wells plugged. Columns (2), (4), and (6) add additional fixed controls for time and location. Columns (3) and (4) restrict the sample to only include the northern region, while columns (5) and (6) restrict the sample to only include the southern region. Clustered at the month level. The set of wells measured are orphaned wells with an unknown owner and missing permit date. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

5 Discussion

The results of tables 3-7 suggest that the number of wells plugged in Ohio counties after the passage of HB225 in 2018 do not change significantly and in some models even decrease. The number of permits issued per county-month increased however, while the probability that permitted wells were drilled increased in the northern set of Ohio border counties. As a result, the net share of wells that were drilled in all probability increased in Ohio after passage of HB225. That is, the proportion of well permits to drilling remained equal while the number of permits increased, thus leading to more drilling occurring in Ohio after 2018. In all sets of regressions the age of wells that were plugged after 2018 was older than those that were plugged prior to 2018 and the passage of the law. The depth of wells plugged varied by model and the results were mixed, but significant. Overall the only conclusive results were that after 2018 in Ohio, more older wells were plugged even though roughly the same number of wells were plugged overall before and after the bill's passage. In addition, more drilling permits were issued after 2018 in Ohio.

While the rise of permits being issued in Ohio after 2018 may or may not have been an unintended consequence of the regulation itself, the lack of any meaningful changes in the number of wells plugged shows that there was no observable offset to the increase in drilling after 2018. The legislation could have been used to create an offset to the expected increase in permit growth, or a way to offset the public bad of environmental damage caused by well drilling by the industry Kotchen (April 2009).[13] The bill failed to address the consequences of more wells being permitted. That is, the data demonstrates a green paradox effect, because the policy may have incentivized resource extraction operators to extract more in the present due to expectations of greater climate policy regulation in the future, as modeled by the Hotelling effect. As Svenn et al. (2015) argue, policies like HB225 disregards supply side behavior that is motivated by the Hotelling rule. Policy makers tend to consider how climate policy will force producers to internalize the environmental costs of fossil fuel usage, but disregard producer response and behavior to these policies. That is, producers will try to maximize their profits in response to climate policy, which leads to greater extraction and drilling in the present due to expectations of greater climate policies being enacted in the future (simple Hotelling rule) as in the case of Ohio HB225. Well drilling may have increased in the present as a result of producer expectations of increased climate policies in the future.

Often a failure to consider producer and consumer behavior can lead to unin tended consequences of regulation.The above considerations suggest an inherent ambiguity in the case of policies designed to limit or control emissions and pollution caused by fossil fuel use or extraction. Cholette and Harrison (2021) discusses the consequences of ambiguity neglect, where policy makers wrongly assume that producers and consumers will act in a predictable way according to the regulations outlined in a given policy. For instance, as Davis (February 2008) showed, in Mexico City a regulation designed to limit automobile emission in the city backfired when consumers adjusted their behaviors in an unexpected way that maximized their own flexibility and ended up increasing pollution

13 Matthew Kotchen discusses the propensity and theory behind private provision of public goods for the purpose of offsetting a public negative externality. In the case of a public good being the environment, which citizens can contribute to by protecting, often times offsets are created, such as in the form of carbon offsets to emissions in a different area (2).

rather than limiting or decreasing it. Producers, like consumers, similarly adjust their behavior based on maximizing returns given the parameters of new regulations as evidenced by the data showing increased permits for drilling in Ohio after 2018.

6 Conclusion

Ohio's House Bill 225, unanimously passed in 2018, was clearly intended to increase the number of orphaned and abandoned wells that would be capped in the state. The results of the study show that older orphaned and abandoned wells were plugged at a higher rate after passage of the bill, but there was no increase in the overall number of wells plugged after 2018. At the same time, this study shows that the number of permits issued per county per month increased by 9. This was not offset by any decrease in the proportion of wells that were drilled. The net effect was therefore an actual increase in the number of wells drilled in the state after 2018, while the rates of plugging remained generally the same. These appear to be unintended consequences of the regulation and they undermine the bill's intended purpose.

Such unintended consequences need to be addressed if the goal is to pass meaningful regulation to limit emissions and to mitigate continued environmental degradation. As Hicks (2019-2020) points out, there are a number of possible solutions to reform the systems through which wells are plugged: adequate reform of the bond system to incorporate financial assurance to cover the costs of decommissioning the wells; collaboration between states to document histories of compliance violations among well operators; and reducing the period for temporary abandonment of wells. However, while these are important and significant solutions to help ensure the tracking and plugging of wells in the Ohio River Basin, none of them address the specific parameters of HB225 which depend on the distribution and use of severance tax funds to plug wells. In part, the bill's failure may be due to the fact that the legislation does not adequately provide implementation strategies as to how the ODNR should go about finding contractors, locating and documenting wells, and organizing a task force within the agency to create a blueprint for allocating resources in a timely manner and addressing supply side behavior. In the future there should be implementation strategies designed to address these unintended consequences.

References

Bang, Sik, and Gary E Hollibaugh Jr. 2021. "Legislative influence on administrative decision making in Pennsylvania's Abandoned and Orphan Well Plugging Program." *Public Administration*, 100(3): 737–758.

Boomhower, Judson. 2019. "Drilling Like There's No Tomorrow: Bankruptcy, Insurance, and Environmental Risk." *American Economic Review*, 109(2): 391–426.

Boutot, Jade, Adam S Peltz, Renee Mcvay, and Mary Kang. 2022. "Documented Orphaned Oil and Gas Wells Across the United States." *Environ. Sci. Technol.*, 56(20): 14228–14236.

Brown, Jason P., Peter Maniloff, and Dale T. Manning. 2020. "Plugging Problems: How States in the Ohio River Basin Can Address Orphan Oil and Gas Wells." *Journal of Environmental Economics and Management*, 103: 1–20.

Brunner, Eric, Ben Hoen, and Joshua Hyman. 1983. "A Theory of Competition among Pressure Groups for Political Influence." *The Quarterly Journal of Economics*, 206(3): 371–400.

Bussewitz, Cathy, and Martha Irvine. July 31, 2021. "Forgotten Oil and Gas Wells Linger, Leaking Toxic Chemicals." *https://apnews.com/article/joe-biden-business-health-environment-and-naturecoronavirus-pandemic-f04ac92a45b4e02392e86b618cc4ff03.*

Cholette, Lor´an, and Sharon G. Harrison. 2021. "Unintended Consequences: Ambiguity Neglect and Policy Ineffectiveness." *Eastern Economic Journal*, 47(2): 206–226.

Davis, Lucas W. February 2008. "The Effect of Driving Restrictions on Air Quality in Mexico City." *Journal of Political Economy*, 116(1): 38–81.

Eubanks, Larry S., and Michael J. Mueller. 1986. "An Economic Analysis of Oklahoma's Oil and Gas Forced Pooling Law." *Natural Resources Journal*, 26(3): 469–491.

Graves, Beth Amy. July 23, 2019. "New Law Speeds Up Plugging of Abandoned Oil, Gas Wells." *https://ofbf.org/2019/07/23/new-law-speeds-plugging-abandoned-oil-gas-wells/.*

Hicks, Connor. 2019-2020. "Plugging Problems: How States in the Ohio River Basin Can Address Orphan Oil and Gas Wells." *Kentucky Journal of Equine, Agriculture, Natural Resources Law*, 12(3): 683–716.

Ho, Jacqueline S, Jhih-Shyang Shih, Lucija A Muehlenbachs, Clayton Munnings, and Alan J Krupnick. 2018. "Managing Environmental Liability: An Evaluation of Bonding Requirements for Oil and Gas Wells in the United States." *Environ. Sci. Technol.*, 52(7): 3908–3916.

Kang, Mary, Adam R Brandt, Zhong Zheng, Jade Boutot, Chantel Yung, Adam S Peltz, and Robert B Jackson. 2021. "Orphaned oil and gas well stimulus—Maximizing economic and environmental benefits." *Elementa: Science of the Anthropocene*, 9(1).

Kotchen, Matthew. April 2009. "Voluntary Provision of Public Goods for Bads: A Theory of Environmental Offsets." *The Economic Journal*, 119(537): 883–899.

Kroszner, Randall S. 2015. "Unintended Consequences: Origins of an Enduring Idea." *Chicago Booth Review*.

Layton, Nicole, and Ginger Sprong. 2022. "Cut and Run: Bonding, Bankruptcies, and the Orphaned-oil-well Crisis." *LSU Journal of Energy Law and Resources*, 10(1): 1–33.

Mallinson, D. J, A Ali, J Guo, and P Robles. 2022. "The Scourge of Orphaned and Abandoned Wells: Leveraging Public-Private-Citizen Collaboration to Solve a Big Problem." *Public Works Management Policy*, 28(1): 33–52.

Nelson, Steve, and Jonathan M Fisk. 2021. "End of the (Pipe)Line? Understanding how States Manage the Risks of Oil and Gas Wells." *RPR Review of Policy Research*, 38(2): 203–221.

Newell, Richard, and Daniel Raimi. 2016. "Oil and Gas Revenue Allocation to Local Governments in Eight States." 1–12.

Newell, Richard G., and Daniel Raimi. 2018. "US State and Local Oil and Gas Revenue Sources and Uses." *Energy Policy*, 112: 12–18.

Ohio General Assembly, Sub. H.B. 225, Ohio Legislative Service Commission. September 28, 2018. *https://www.legislature.ohio.gov.*

Peltzman, Sam. 1976. "Toward a More General Theory of Regulation." *The Journal of Law and Economics*, 19: 211–240.

Raimi, Daniel, Alan J Krupnick, Jhih-Shyang Shah, and Alexandra Thompson. 2021. "Decommissioning Orphaned and Abandoned Oil and Gas Wells: New Estimates and Cost Drivers." *Policy Analysis*, 55(15): 10224–10230.

Schuwerk, Robert, and Greg Rodgers. 2020. "Billion Dollar Orphans: Why millions of oil and gas wells could become wards of the state."
Simmers, Rick. January 16, 2018. "House Energy and Natural Resource Committee Interested Party Testimony on House Bill 225." *https://www.legislature.ohio.gov/legislation/132/hb225/committee.*

Stewart, Thomas E. February 28, 2018. "Proponent Testimony Regarding House Bill 225 Idle and Orphan Oil and Gas Wells." *https://www.legislature.ohio.gov/legislation/132/hb225/committee.*

Stigler, George J. 1971. "A Theory of Market Regulation." *The Bell Journal of Economics and Management Science*, 2: 3–21.

Svenn, Jensen, Kristina Mohlin, Karen Pittel, and Thomas Sterner. 2015. "An Introduction to the Green Paradox: The Unintended Consequences of Climate Policies." *Review of Environmental Economics and Policy*, 9(2): 246–265.

The API Number Standard: An Identifier for Petroleum Industry Wells in the USA. 2013. *https://dl.ppdm.org/dl/836.*

Wagoner, Rachel. August 18, 2022. "Audit Finds Ohio orphan well program falls short." *https://www.farmanddairy.com/news/audit-finds-ohio-orphan-well-program-falls short/730991.*

A Appendix

This section presents robustness checks beyond the extensions presented in the body of the paper. It also presents graphical interpretations of the data through observed means, and a visualization of an event study where the base group is the year prior to the passage of HB225. The event study finds similar outcomes to the difference-in-differences model presented in the paper. Permits are issued at a greater rate following the passage of HB225, and this effect stays consistent over the four year period following the passage of the law. There is no observable shift in the rate at which wells are plugged, and there is a dip in the probability that a drilled well is plugged.

A.I Appendix Tables

Tables A.1 and A.2 are extensions of table 3 that measure permit depth and age separately as dependent variables. Table A.1 reports the effect of HB225 on the average proposed total depth (PTD) issued on permits from September 2014-2022. All specifications show a negative effect of treatment on proposed total depth, indicating that shallower wells were plugged after the passage of HB225. However, Ohio wells had much higher PTD's than Pennsylvania, indicating that projects issued on permits in Ohio were generally large and costly. Table A.2 presents the effect of HB225 on the average permit age (age to approval) of wells. Results show that the bill did not have an effect on the amount of time it took for permits to be approved. On average, permits in Ohio required significantly more days before approval compared to wells in Pennsylvania.

Table A.1: Proposed Total Depth(ft) Per County-Month Using Permit Dataset

	All Border Counties		Northern Region		Southern Region	
	(1)	(2)	(3)	(4)	(5)	(6)
OH County Post	-1569.51***	-1572.78***	-471.32*	-472.20*	-2646.75***	-2646.75***
	(494.57)	(496.23)	(272.52)	(271.54)	(917.94)	(921.17)
Ohio County	5757.39***	5757.76***	1619.68***	1618.30***	8615.69***	8615.69***
	(287.49)	(288.18)	(200.46)	(201.03)	(520.04)	(521.87)
Post Period	-200.51***	855.45***	-155.25***	370.61**	-245.76***	1355.26***
	(32.03)	(308.69)	(34.43)	(153.52)	(50.64)	(542.89)
Near Marcellus Shale Gas Play	Yes	Yes	Yes	Yes	No	No
during Biden Controls	No	Yes	No	Yes	No	Yes
Oil And Gas Price Control	No	Yes	No	Yes	No	Yes
Pandemic (2020) Control	No	Yes	No	Yes	No	Yes
n =	1152	1152	576	576	576	576
R^2	0.287	0.301	0.177	0.205	0.285	0.306

Notes: This table is an extension to table 3 and includes the regressions of average proposed total depth (as listed on the permit) per county-month on the difference-in-differences treatment variables. Columns (1), (3), and (5) do not include price and time controls while columns (2), (4), and (6) include these controls. Clustered at the month level. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

Table A.2: Permit Age Per County-Month (days) Using Permit Dataset

	All Border Counties		Northern Region		Southern Region	
	(1)	(2)	(3)	(4)	(5)	(6)
OH County Post	-153.11	-152.63	-242.10	-240.86	-65.14	-65.14
	(164.65)	(165.11)	(327.84)	(329.66)	(39.77)	(39.91)
Ohio County	376.87***	376.60***	764.85***	763.69***	83.72***	83.72***
	(100.52)	(100.68)	(222.53)	(223.20)	(20.91)	(20.99)
Post Period	14.10	32.86	-1.00	20.68	29.20	44.88
	(11.45)	(76.71)	(1.87)	(153.80)	(22.90)	(26.26)
Near Marcellus Shale Gas Play	Yes	Yes	Yes	Yes	No	No
during Biden Control	No	Yes	No	Yes	No	Yes
Oil And Gas Price Controls	No	Yes	No	Yes	No	Yes
Pandemic (2020) Control	No	Yes	No	Yes	No	Yes
n =	1152	1152	576	576	576	576
R^2	0.028	0.030	0.177	0.035	0.017	0.019

Notes: This table is an extension to table 3 and includes the regressions of average permit age per county-month on the difference-in-differences treatment variables. Columns (1), (3), and (5) do not include price and time controls while columns (2), (4), and (6) include these controls. Clustered at the month level. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

Tables A.3 and A.4 present sets of regressions that limit the post-period in order to exclude pandemic years. All specifications end at February 2020, corresponding to the start of the pandemic. Columns (1), (4), and (7) include the full pre-period. Columns (2), (5), and (8) restrict the pre-period to September 2016 - February 2020, and columns (3), (6), and (9) further restrict the pre-period to include identical months before and after the passage of HB225 (April 2017 - February 2020). Interestingly, in Table A.3, there appears to be an increasing treatment effect as the time period becomes smaller. In the southern region, the treatment effect is quite substantial. In the first entry of column (9), over 8 permits more were issued per county-month following HB225 when the observed time period is restricted to roughly 3 years. Table A.4 shows some significant changes based on specification, but no positive effects of HB225 on the number of wells plugged per county-month. The age and depth results from table A.4 show similar effects to table 4 presented in the body of the paper. Shallower wells appear to be plugged at a greater rate after treatment, although only significantly in specifications (1) and (2). Well age appears to be positively correlated with treatment in all border counties, but this effect is due to the positive effect from the southern region only.

Table A.3: Permits Issued Per County-Month Using Permit Dataset: Restricting Months After Pandemic

	All Border Counties			Northern Region			Southern Region		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OH County Post	2.409***	2.542*	4.537***	0.549**	0.0665	0.187	4.205**	4.913*	8.754***
	(0.866)	(1.315)	(1.522)	(0.272)	(0.289)	(0.312)	(1.649)	(2.561)	(3.015)
Ohio County	-3.498***	-3.649***	-5.756***	0.410**	0.712***	0.600***	-9.597***	-10.306***	-14.146***
	(0.665)	(1.169)	(1.356)	(0.163)	(0.181)	(0.203)	(1.373)	(2.389)	(2.869)
Post Period	-3.112***	-3.645***	-4.971***	-0.509***	-0.0718	-0.114	-5.716***	-7.223***	-9.837***
	(0.767)	(1.065)	(1.349)	(0.140)	(0.101)	(0.126)	(1.506)	(2.137)	(2.716)
Near Marcellus Shale Gas Play	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Oil And Gas Price Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
n =	780	492	396	390	246	198	390	246	198
R^2	0.211	0.213	0.231	0.054	0.110	0.102	0.186	0.221	0.315

Notes: This table reports the regressions for permits issued per county-month on the difference-in-differences treatment variables for time periods that exclude the pandemic. Column (1) reports the coefficients for the treatment variables using the time frame from September 2014 to February 2020 (the start of the pandemic). Column (2) further restricts the time period from September 2016 to February 2020, and column (3) uses the time period from September 2017 to February 2020. Columns (4) - (6) restrict the sample to the northern region counties, while columns (7) - (9) restrict the sample to southern region counties. Clustered at the month level. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

Table A.4: Plugged Wells Per County-Month Using Permit Dataset: Restricting Months After Pandemic

	All Border Counties			Northern Region			Southern Region		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OH County Post	-0.195	-0.459*	-0.482***	-0.021	-0.182	-0.079	-0.459	-0.770*	-0.969*
	(0.238)	(0.263)	(0.310)	(0.250)	(0.255)	(0.266)	(0.379)	(0.432)	(0.482)
Well Depth * OHCountyPost	-0.208***	-0.125*	-0.463	0.014	0.077	0.091	-0.210	-0.336	-0.463
	(0.078)	(0.065)	(0.067)	(0.102)	(0.103)	(0.117)	(0.128)	(0.249)	(0.331)
Well Age * OHCountyPost	0.029***	0.029***	0.030***	-0.004	0.003	0.0013	0.036***	0.037***	0.039***
	(0.0082)	(0.0079)	(0.0073)	(0.015)	(0.015)	(0.015)	(0.0091)	(0.0092)	(0.0087)
Average Well Depth (Thousands of ft)	0.324***	0.234**	0.227***	0.066	-0.014	-0.049	0.332***	0.450*	0.563*
	(0.051)	(0.038)	(0.043)	(0.055)	(0.059)	(0.077)	(0.116)	(0.242)	(0.327)
Average Well Age (yrs)	0.016***	0.015***	0.014***	0.044***	0.034***	0.033***	0.011***	0.012***	0.011***
	(0.0023)	(0.0023)	(0.0021)	(0.0077)	(0.0078)	(0.0089)	(0.0022)	(0.0027)	(0.0027)
Ohio County	-1.235***	-0.967***	-0.924****	-0.260**	-0.124	-0.167	-2.033***	-1.705***	-1.504***
	(0.129)	(0.165)	(0.216)	(0.117)	(0.129)	(0.177)	(0.187)	(0.271)	(0.331)
Post Period	0.162	0.288	0.184	0.021	0.142	-0.011	0.348	0.483	0.497
	(0.244)	(0.268)	(0.297)	(0.208)	(0.216)	(0.244)	(0.405)	(0.447)	(0.477)
Marcellus Gas Play Control	Yes	Yes	Yes	No	No	No	No	No	No
Oil and Gas Price Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Combined Oil and Gas Production Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
n =	780	492	396	390	246	198	390	246	198
R^2	0.339	0.363	0.362	0.371	0.419	0.469	0.344	0.346	0.345

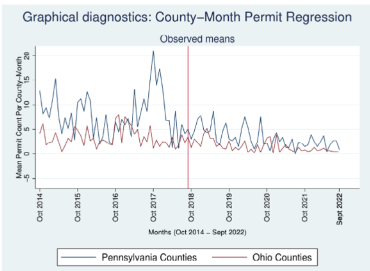
Notes: This table reports the regressions for plugged wells per county-month on the difference-in-differences treatment variables and a set of control variables for time periods that exclude the pandemic. Column (1) reports the coefficients for the treatment variables using the time frame from September 2014 to February 2020 (the start of the pandemic). Column (2) further restricts the time period from September 2016 to February 2020, and column (3) uses the time period from September 2017 to February 2020. Columns (4) - (6) restrict the sample to the northern region counties, while columns (7) - (9) restrict the sample to southern region counties. Clustered at the month level. Significantly different than zero at 99 (***), 95 (**), 90 (*) percent confidence.

A.II Appendix Figures

Figures used in this paper are designed to provide visualization of the difference-in differences model and an event study that analyzes the effects of HB225 over time. A.1 presents the observed means for the permits issued and wells plugged per county-month. In both panels, the observed mean over time appears to be higher in Pennsylvania counties, but in panel A, there is an observable decline in permits issued per county-month on average. Permits issued in Ohio remain fairly consistent over time and do not decline during the pandemic. The trends in wells plugged per county-month in panel B do not appear to shift significantly in either state.

Figure A.1: Observed Means and Trends For County-Month Comparisons

Panel A: Permitted Wells



Panel B: Plugged Wells

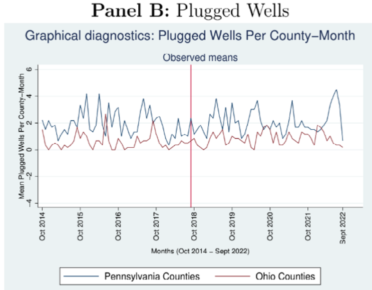
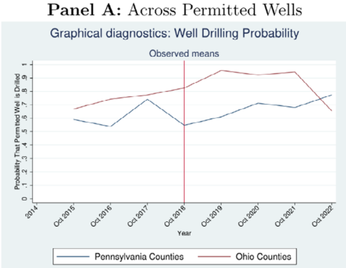


Figure A.2 presents the observed means on a yearly level for the probabilities measured in the LPM models. Panel A shows an upward trend in the well drilling probability in Ohio. In panel B, trends in both states appear to be positive over time. Figure A.3 presents the observed yearly trends in the probability that orphaned wells are plugged. There is a clear upward trend in the probability that a plugged well is orphaned in Ohio, and a slight upward trend in Pennsylvania.

Figure A.2: Observed Means and Trends for LPM Comparisons

Panel A: Across Permitted Wells



Panel B: Across Drilled Wells

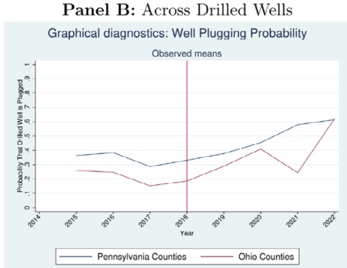
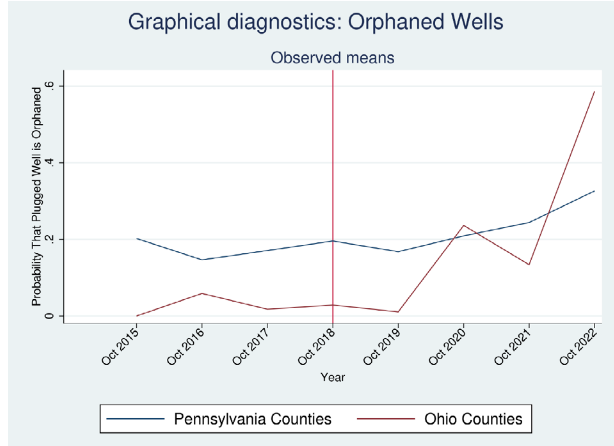


Figure A.3: Graphical interpretation: Table 8



The next set of figures presents the results of the event study conducted for each model presented in the paper. These event studies model the treatment variable over time to see if treatment varies by year. The base period is the 12 month period prior to the passage of the bill on September 28, 2018. Each figure corresponds to an event study conducted with the set of controls listed from the matching table. Figure A.4 shows the trend in permits issued per county-month due to the treatment effect of the bill. The graph clearly shows a positive effect of HB225 on permits issued, and this is sustained each year in the post-treatment phase. Figure A.5 shows the trend in wells plugged per county-month due to the treatment effect of the bill. This figure shows no post-trend deviation from the control time period, although the confidence intervals are quite large.

Figure A.4: Event Study: Table 3

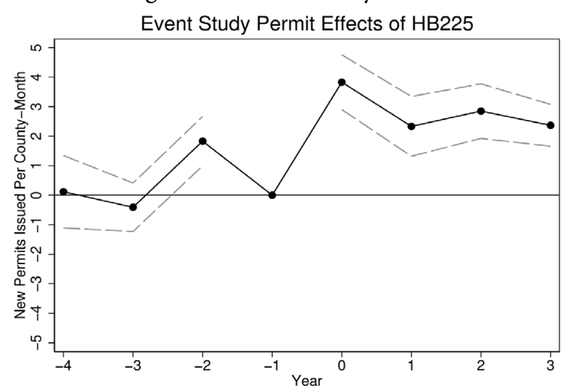


Figure A.5 Event Study: Table 4

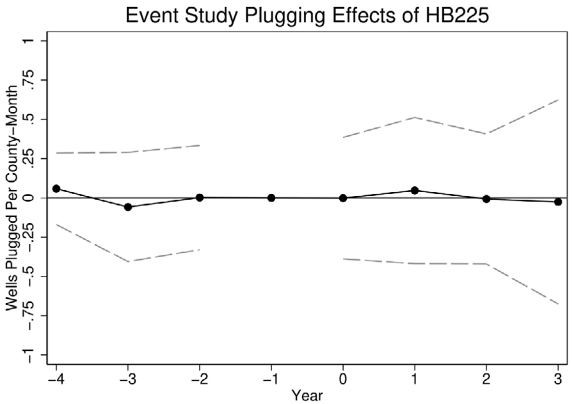


Figure A.6 shows no change in the treatment variable until the last year of the event study, where the share of permitted wells that are drilled falls by 35% (between a confidence interval of -16% and -54%). There also appears to be a pre-trend where treatment occurs before the passage of the law, but the confidence intervals do not exclude 0. These results are slightly different from the results presented in table 5 from the body of the paper. In particular, the decrease in probability in the 4'th year is not captured in the difference in differences model for table 5.

Figure A.6: Event Study: Table 5

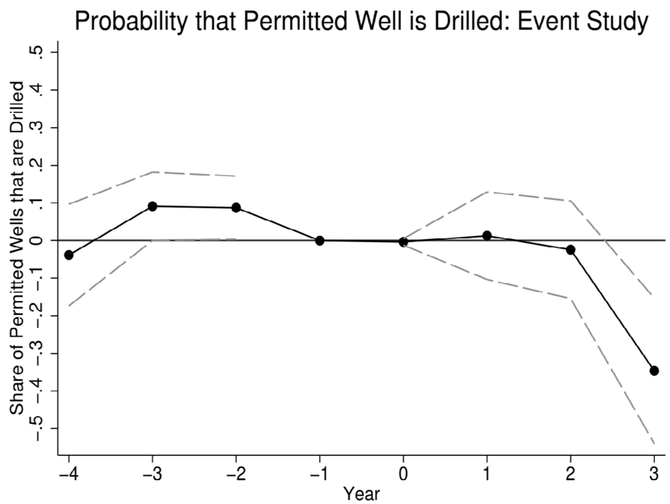


Figure A.7 shows a sustained decrease in the share of drilled wells that were plugged in Ohio following HB225. There is also a slight positive pre-trend which does not exclude 0. These results match the results presented in the linear probability models for table 6 from the body of the paper. Figure A.8 presents the treatment on the share of orphaned wells plugged from table 8. There is a positive effect on the share of orphan wells plugged over time, but the confidence intervals increase as the treatment moves farther from the event.

Figure A.7: Event Study: Table 6

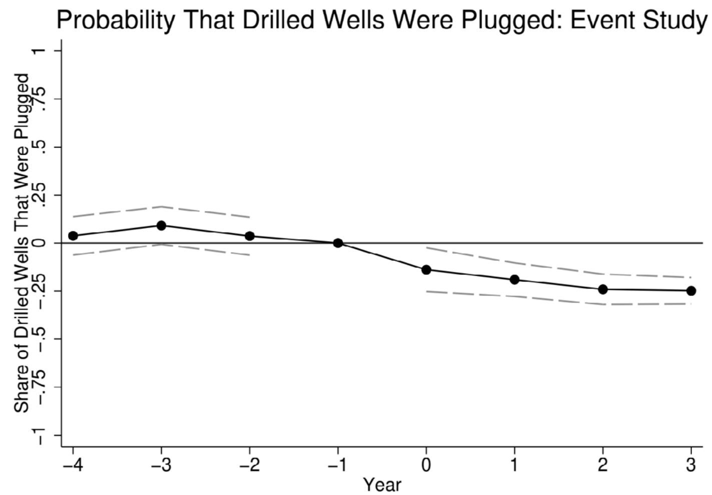
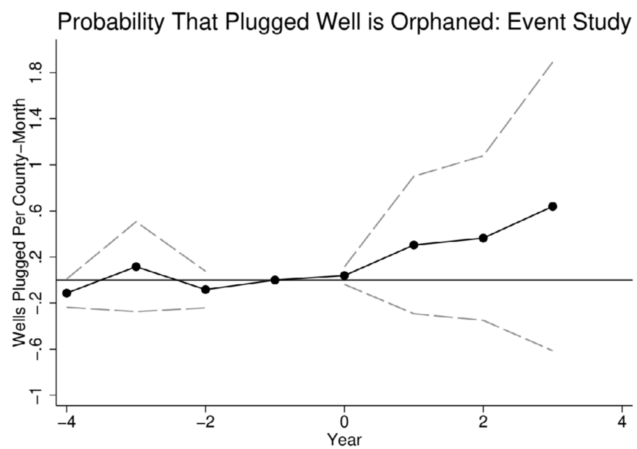


Figure A.8: Event Study: Table 8



The Effect of an Extension in Potential Unemployment Insurance Benefits Duration on Unemployment Duration: Evidence from Italy

September 16, 2023

Abstract

I estimate the causal effect of an 'extension' in potential Unemployment Insurance (UI) benefits duration on the duration of unemployment (UD). I exploit two age discontinuities in the eligibility for longer benefits durations of the Italian UI program ASpI in the years 2014 and 2015. I find that, on average, a one-week unemployment benefits extension increases UD by between 0.02 and 0.65 weeks. I also conduct discrete hazard and survival analysis to check for a 'spike' at benefits exhaustion and characterize the path of the re-employment escape rate, but observe neither an increase in the job-finding rate as benefits expire nor the U-shaped pattern fleshed out by previous studies. While the direction of causality of the regression discontinuity estimate is consistent with the consensus in the literature, confidence intervals are large. Furthermore, I cannot rule out a non-finding at the 95% confidence level for my main estimates. This and our negative findings on the path of the re-employment hazard are due to the severely limited nature of the publicly available state-level Italian labor-force survey data used in the analysis.

Keywords: Unemployment Duration, Unemployment Insurance Benefits, Discrete Duration Analysis

JEL Codes: C41, J64, J65

Content

1 Introduction

The empirical analysis of the effect of an extension in potential UI benefits duration on UD is salient because it allows for the validation of labor-search models. Early works in labor search theory Mortensen (1977); van den Berg (1990) have posited that the employment escape rate increases around and until benefits exhaustion. Hence, extending potential UI benefits duration brings forth a reduction of employment hazard, thereby leading to longer UD. Such a clear and testable hypothesis sparked a sizeable empirical literature trying to estimate the elasticity of UD with respect to UI benefits duration extensions, as well as the path of the employment hazard through the unemployment spell. While this stream of research has been able to establish a strong consensus around the elasticity's positive direction, there is still variability in its actual magnitude. Furthermore, there is growing evidence that the empirical hazard does not fully match the path predicted by the early theoretical works. In this regard, the Italian unemployment insurance program in place in the 2014-2015 period offers fertile ground for an empirical analysis of how extended UI benefits affect UD, and of the path of the employment hazard.

Over the 2010s Italy changed its unemployment benefit eligibility cohort and benefit amounts twice. In the pre-2013 period Italy had three UI benefit programs called *Indennità di disoccupazione, Indennità di disoccupazione a requisiti ridotti*, and *Mobilità*. The 2012 LEGGE 28 giugno 2012, n. 92, in effect on January 1st 2013, supplanted the *Indennità di disoccupazione*, and the *Indennità di disoccupazione a requisiti ridotti*, and introduced the so-called *Assicurazione sociale per l'impiego (ASpI)*, and, with the intent of enlarging the eligibility cohort even further, the so-called mini-ASpI. A further homogenization happened in 2015 with the *D.Lgs. 4 marzo 2015, n. 22*. The latter unified ASpI and mini-ASpI in what became known as the *Nuova assicurazione sociale per l'impiego* (NASpI). A more in-depth description of the programs' specifics (eligibility requirements, benefit amounts, variation of benefit amounts as a function of benefits duration, etc.), is in order in Section 8.

I analyze the difference in UI benefit lengths as a function of age. Both the *Indennità di disoccupazione*, the *Mobilità*, and ASpI assign benefits for a shorter time to recipients under a certain age threshold, which is different for each program, and for some programs different every year (see Giorgi (2018) and Section 8). By exploiting this discontinuity I perform an RD design regressing age on UDs. I thus compare the effect of extended UI benefits on UDs for ASpI recipients in 2014 and 2015. Due to data limitations that will be discussed further in the paper I am not able to analyze the effect of extended UI benefits on UDs for what concerns the *Indennità di disoccupazione* and the *Mobilità* programs. Furthermore, the *Indennità di disoccupazione a requisiti ridotti* and the NASpI programs do not come with age discontinuities in UI benefits duration. Hence, they too are outside of the scope of our analysis.

I use publicly available ISTAT survey data called *Rilevazione Continua sulle Forze di Lavoro* (RCFL) to carry out the analysis. The latter is standardized according to the REG ULATION (EU) 2019/1700, and corresponds to similar Labor Force Surveys in other European countries. It uses a continuous sampling strategy of over 250 thousand Italian families, which are interviewed four times in the time span of 15 months. Since we know whether respondents are employed at each survey time, had a previous unemployment spell, and their age, I have sufficient information to carry out our RD design, as well as to characterize hazard and survival functions.

In Section 2 I delve into the main theoretical frameworks and the ensuing empirical literature concerning the effect of UI benefits on UD, in Section 3 I further expand on the relevance of our study, in Section 4 I describe the data used for our analysis, our sampling strategy, and descriptive statistics, in Section 5 I present our main specifications and empirical strategy, in Section 7 I carry out robustness checks and sensitivity analysis for the main specification, and in Section 8 I discuss my results and conclude.

2 Theoretical Framework and Existing Literature

Theoretical labor-search models represent a substantial and growing body of research, whose inception is often attributed to Stigler (1961) (see Mortensen (1986); Rogerson et al. (2005)). This section does not aim to present a detailed historical compound of such models but instead focuses on salient works for empirical analysis. Furthermore, it offers a detailed account of existing estimates of the effect of UI benefit extensions on UD, as well as a review of the behavior of re-employment hazard around benefits exhaustion.

2.1 Labor Search Theory and Unemployment Insurance Benefits

In Mortensen (1977) labor search model unemployed workers choose their search intensity s, whose cost is considered the value of the leisure forgone, to maximize the future sequence of consumption and leisure pairs. They sample sequentially from a known and stationary cumulative wage distribution $F(w)$, and their optimal strategy is to accept the first obtained offer greater than their reservation wage. Finally, qualified workers obtain UI benefits for a finite duration. In such a setting, the model predicts that the escape rate from unemployment q increases with the realized unemployment duration. As Meyer (1990) points out, such a prediction is easily grasped if we understand that the escape rate is proportional to $s[1-F(w)]$, where w is the reservation wage, and $1-F(w)$ is the probability that a given job offer is acceptable. Since the reservation wage w decreases as benefits approach exhaustion, given that the value of being unemployed decreases, and since the marginal benefit of search increases when the value of unemployment declines, increasing search intensity s, the escape rate q increases until benefits exhaustion. The model predicts a subsequent constant hazard thereafter since the conditions of the unemployed worker remain unchanged. Instead, for what concerns workers not eligible, or having exhausted benefits, q is thought of as being independent of the realized unemployment duration. This can be understood intuitively by thinking that if income and leisure are complements, then a worker receiving UI benefits will have a higher marginal utility from leisure. Finally, given that q increases until benefits exhaustion, an extension in the duration of UI benefits will increase the value of unemployment contingent on receiving the longer benefits duration, hence decreasing q and lengthening the UD.

Moffitt and Nicholson (1982) adopt a static approach in the tradition of consumption leisure trade-off models, rather than search-theoretic ones. Individuals maximize utility as a function of income and unemployment, whereby unemployment increases utility because it provides leisure and productive job search time. In such a context the opportunity cost of being unemployed is the forgone wage W workers could have earned, had they worked instead. However, if the unemployed worker is perceiving UI benefits the opportunity cost of not working is reduced to $W(1-r)$ where r is the replacement rate, namely the ratio of perceived benefits B to previous income $(r = \frac{B}{W})$. Thus, the budget constraint of a worker receiving benefits will be kinked, whereby the first segment will have a slope of negative $W(1-r)$ and the second segment a slope of negative W, with $|W(1-r)| < |W|$. Moffitt and Nicholson (1982) shows that in such a framework if the potential benefit duration U^* were to increase, then a worker who would exit unemployment before benefits exhaustion would not be affected, while a worker who would be placed at the kink point or in the second segment of the budget constraint would see an increase in their UD. Unlike Mortensen (1977), Moffitt and Nicholson (1982) does not provide a formal argument for an increasing hazard around benefits exhaustion, but it posits that if we consider a continuous distribution of preferences, workers with $|W(1-r)| < |MRS| < |W|$ will place at the kink point. Hence, heuristically, re-employment hazard will rise at benefits exhaustion.

Card et al. (2007a) uses a Lentz and Tranæs (2005) setting, whereby individuals maximize a separable utility function in consumption c_t, which comes in positive, and search effort s_t, which comes in negative $u(c_t)-\psi(s_t)$. This set of models incorporates savings and wealth in the worker's decision considering wages as exogenously determined, as opposed to Mortensen (1977). The unemployed worker chooses s_t at the beginning of period t to maximize the expected utility over the two states of the world $s_t V_t(A_t) + (1-s_t) U_t(A_t) - \psi(s_t)$. Here A_t represents the beginning of period assets, $V_t(.)$ represents the value function conditional on finding employment at the beginning of period t, and $U_t(.)$ represents the value function conditional to still being unemployed at the beginning of period t. Intuitively s_t is chosen to equate the marginal value of search effort and the marginal cost of search effort given by the difference between the optimized value of employment and unemployment. Card et al. (2007a) shows that under these conditions the partial derivative of search effort with respect to future benefits $\partial s^*_t/\partial b_{t+j}$ is weakly negative, meaning that s_t decreases in future benefits. Hence, even in the case where saving decisions and wealth are taken into account an extended benefit duration will bring about an extended UD, since it reduces search effort over the whole period.

One of the most interesting and recent theoretical contributions to labor search theory is possibly DellaVigna et al. (2017). It uses the framework laid out by Lentz and Tranæs (2005) to account for workers' so-called reference-dependent preferences understood in the context of Kahneman and Tversky (1979) prospect theory. Agents with reference-dependent preferences over utility from consumption are loss averse below an average income reference point $r_t = \frac{1}{N+1}\sum_{k=t-N}^{t} y_k$, where N represents all previous periods, and t represents the current period. Thus, agents have a backward-looking reference point as in Bowman et al. (1999). Their utility is a function of consumption c_t conditional on the reference point r_t, based on Kőszegi and Rabin (2006) specification

$$u(c_t \mid r_t) = \begin{cases} c_t + \eta[(c_t) - (r_t)] & \text{if} \quad c_t \geq r_t \\ c_t + \eta\lambda[(c_t) - (r_t)] & \text{if} \quad c_t < r_t \end{cases}$$

Here $v(c_t)$ is utility from consumption in period t, and $v(c_t) - v(r_t)$ is gain-loss utility. If consumption is greater than the reference point $c_t \geq r_t$, the individual receives gain utility $v\ (ct) - v\ (rt) > 0$ with weight η. If consumption is smaller than the reference point $c_t < r_t$, then the individual has loss utility $v\ (ct) - v\ (rt) < 0$ with weight $\lambda\eta$, where $\lambda \geq 1$ represents loss aversion, namely that marginal utility is higher for losses than for gains. In such a framework DellaVigna et al. (2017) predicts that at the beginning of the UI spell worker's reference point is high, hence unemployment is painful and search will be intensive. As the unemployment spell progresses, the reference point shifts to a lower level since benefits are lower than previous income and search intensity slows down. As workers approach benefits exhaustion they anticipate the future loss in consumption and search more intensively again. Finally, after benefits are exhausted the reference point shifts lower once again, and search intensity decreases. Since search intensity is proportional to the escape rate as in Mortensen (1977), the escape rate will follow the just delineated path. Hence, DellaVigna et al. (2017) prediction differs from Mortensen (1977) in two respects: (1) the escape rate falls during the UI benefits spell has a U-shape, as opposed to being non-decreasing; (2) the escape rate falls after benefits exhaustion, as opposed to staying constant.

The just presented theory unanimously predicts an increase in UD as a consequence of UI benefits extension. Nevertheless, DellaVigna et al. (2017) posits its model against the predictions of Mortensen (1977). Said that I will now present evidence of such effects. I will come back to the models' predictions in Section 3 where I will state the hypotheses I will set out to test in the remainder of the paper.

2.2 The Effect of an 'Extension' in Potential Unemployment In surance Benefits Duration on Unemployment Duration

2.2.1 Evidence from the United States

Early estimates of prolonged UI benefit duration on UD come from the US Moffitt and Nicholson (1982); Moffitt (1985); Katz and Meyer (1990b). These studies find significant effects of UI benefit extensions on UD, albeit small. Moffitt and Nicholson (1982); Moffitt (1985) use variation in potential benefits duration resulting from the 1971 Extended Benefits program (13 additional weeks of potential benefits duration during cyclical downturns) and the Federal Supplemental Benefits Program triggered by the 1975-75 and 1981-92 recessions (26 to 39 weeks) for identification. They find that a 1-week increase in potential UI benefit duration would lengthen UD by 0.1 weeks and 0.16 weeks respectively. Katz and Meyer (1990a) exploits cross-sectional variation in benefits length for 12 US states and finds that a 1-week increase in potential UI benefit duration would lengthen UD by about 0.16-0.2 weeks. This early evidence seems to validate the predictions of the presented labor-search theory. Card and Levine (2000) represents an early attempt to control labor market conditions. It examines an exogenous (politically determined) 6 months-long 13 weeks lengthening of potential benefits duration in New Jersey in 1996. It obtains a week-per-week estimate of about 0.076, which is markedly lower as compared to the earlier studies.

2.2.2 Evidence from Europe

Hunt (1995); Winter-Ebmer (1998); Lalive and Zweimüller (2004); van Ours and Vodopivec (2006); Lalive (2008); Schmieder et al. (2016); Nekoei and Weber (2017) are instead notable studies of the effect of an extension in potential benefits duration on UD in the European context. Hunt (1995) uses a 1980s German policy change lengthening potential benefits duration for unemployed aged 41 and older and finds an estimate consistent to but somewhat smaller than Moffitt (1985). Winter-Ebmer (1998) exploits a quasi-experimental design brought about by an Austrian 1988 law change extending potential benefit duration for elderly workers in certain regions only. It estimates that a 1-week potential UI benefits duration extension increases UD by 0.03 weeks. Furthermore, it argues that since the policy change only affected a minority of Austrian workers, general equilibrium arguments would push down the estimate even more. van Ours and Vodopivec (2006) exploited a DiD approach to study a natural experiment in Slovenia whereby the potential UI benefits duration was decreased by half for certain groups and remained constant for others, finding that for every week cut from potential benefits UD decreased by about 0.2 weeks. Nevertheless, Card et al. (2007a) point out that as reported by Vodopivec (2016) Slovenia has a high proportion of workers engaged in the informal sector who strategically wait until the expiration of UI benefits, hence van Ours and Vodopivec (2006) might be over-estimating the true effect. Further estimates from Austria are provided by Lalive and Zweimüller (2004); Lalive (2008), who investigate an extension of penitential benefits duration as a function of age (August 1989) and region (June 1988) to a maximum of 209 weeks for elder workers (50 or older). The latter potential benefits extension was part of an endogenous program aimed at providing relief for the crisis-ridden steel sector in place from 1988 to 1993. Lalive and Zweimüller (2004) accounts for such endogeneity by (1) focusing the analysis on non-steel sectors, (2) focusing on a subset of regions that lost eligibility to the program to account for spillovers, and (3) using a further 1991 normative change requiring that not only the claimant had to be a resident in one of the program's selected regions, but also the employer. They show that in case (1) the week-per-week effect of an extension in potential benefits duration amounts to 0.09, while in cases (2) and (3) to 0.055. Lalive (2008) adopts, instead, an RD design exploiting the age and geographical discontinuities in potential benefits duration of the Austrian program, finding a magnitude of 0.09 for men and 0.32 for women. The heterogeneous effect of the extension in potential benefits duration with respect to gender is accounted for by the availability of early retirement pathways for women significantly earlier than for men (5 years earlier). More recently Schmieder et al. (2016); Nekoei and Weber (2017) also use RD identification strategies and find week-per-week effects of about 0.15 in Germany (cf. Hunt (1995)) and close to 0 in Austria (cf. Winter-Ebmer (1998); Lalive and Zweimüller (2004); Lalive (2008)).

2.2.3 Evidence from Italy

Finally, Rosolia and Sestito (2012); Scrutinio (2019); Citino et al. (2019) are the main existing studies of the disincentive effect of an extension in potential benefits duration in the Italian context. Rosolia and Sestito (2012) analyzes a 3 months extension (from 6 to 9) in the potential benefits duration of 50 years or older workers put into law in 2001 pertaining

to the *Indennità di disoccupazione* program. It establishes a week-per-week effect between 1.53 and 2.3. These estimates are large because Rosolia and Sestito (2012) adopts benefits length as the outcome variable, rather than non-employment spells. As pointed out by Mortensen (1977) in its review of early empirical studies on the matter, such practice can lead to an upward truncation bias. Mortensen (1977) shows that a rise in potential benefits duration increases the escape rate of an ineligible worker if he will be eligible for benefits in the future since the indirect utility of employment rises because of the future reduction in the cost of being laid off. Furthermore, if the probability of exhausting benefits decreases in potential benefits duration, then using benefits length as an outcome variable does not capture the increase in the escape rate of exhaustees. Scrutinio (2019) also analyzes the *Italian Indennità di disoccupazione* program, but in the 2009 and 2012 period. During the time frame under consideration, the age discontinuity consisted of 4 more weeks in potential benefits duration for workers 50 years or older. They find a week-per-week effect between 0.358 and 0.46. Finally, Citino et al. (2019) considers the same setting of Scrutinio (2019), but uses a 'donut' RD design to correct for the sizeable manipulation in the exact timing of lay-offs around the age-at-layoff threshold. It finds that the average population week-per-week effect is between 0.12 and 0.55, slightly smaller than Scrutinio (2019). Furthermore, implementing bunching estimators *à la* Saez (2010); Chetty et al. (2011); Kleven and Waseem (2013) they quantify manipulation to be 15.8% of all lay-offs in the missing density region before the 50 years threshold, and 20.3% of all lay-offs in the excess density region after the 50 years threshold.

2.3 Evidence of a 'Spike' at Unemployment Insurance Benefits Exhaustion

Evidence of the effect of UI benefits on UD has also come from studying the rate of leaving UI roles in the weeks before benefit exhaustion.

Meyer (1990); Katz and Meyer (1990a,b); Moffitt (1985) find a sharp increase in the escape rate from unemployment both through recalls and new job acceptances for UI re cipients around the time of benefits exhaustion in the US context. They also show that increases are not apparent at similar points of spell duration for non-recipients. However, the 'spike' in the hazard rate of UI benefit recipients at benefits exhaustion has been brought into question by Card et al. (2007b) studying the case of Austria. It argues that the increase in hazard rates of UI benefit recip ients at benefits exhaustion is largely due to the way we measure UDs. Spikes are generally smaller when the spells are measured by the time to next job than when they are defined by the time spent in the unemployment system Card et al. (2007b). Furthermore, Card et al. (2007b) find that only one percent of jobless spells have an ending date that is manipulated to coincide with the expiration of benefits in Austrian data. van Ours and Vodopivec (2006) do identify a clear spike at benefits exhaustion. It also finds that the rate of job finding after benefit expiration was substantially higher than be fore, hence lending empirical backing to Mortensen (1977). Finally, DellaVigna et al. (2017) develops formal goodness of fit measures to test its reference-dependent model against the standard model (Meyer (1990)) in the context of the Hungarian UI system. It shows that the reference-dependent model fits the data better than the standard model, especially after benefits exhaustion, whereby the empirical hazard monotonically decreases as opposed to the standard model. It also confirms the U-shaped path of the empirical hazard during benefits reception.

Hence, while there seems to be significant evidence that a 'spike' at benefits exhaustion is present, the prediction that hazard is monotonically increasing during benefits reception and that it is constant after exhaustion is questioned.

3 Contribution and Testable Hypotheses

My paper contributes to the existing literature on the effect of extended UI benefits on UD in the context of an unstudied Italian UI regime and is relevant in three distinct and unrelated regards.

First, while there is an empirical consensus around the positive sign of the effect of extended UI benefits on UD, as shown in Section 2 studies conducted in comparable labor markets and business cycle conditions obtain different magnitudes of the effect. Furthermore, Atkinson et al. (1984) warns us about the possible risks of misspecification and of the quality of the micro-data used for the analysis of our topic. Hence, it is important to provide additional estimates of such an effect. Moreover, economists question whether the job search models presented in Section 2 actually capture all the relevant determinants of the effect of UI benefits duration on UD. For instance, Rogers (1998) introduces the importance of information and expectation about UI benefits of unemployed individuals in determining UD outcomes, while Ben-Horim and Zuckerman

(1987) shows that the presence of financially constrained individuals could actually reverse the positive sign of the effect of an extension in potential benefits duration since those agents could be using the benefits to intensify search efforts.

Second, as shown in 2 there are very few studies that look at the effect of UI benefits extensions on UD in Italy Rosolia and Sestito (2012); Giorgi (2018); Scrutinio (2019); Citino et al. (2019). This could be due to the fact that the most extensive data source to study such phenomena is confidential data collected by the *Italian Social Security National Institute* (INPS) and hard to obtain. On top of the relevance of carrying out such a study in Italy, our investigation is important because it concerns itself with normative changes that few economists have investigated, and if they have (see Giorgi (2018)) they have looked at different aspects of the reforms (coverage, revenue effects, take-up rates, etc.), or different years (see Scrutinio (2019); Citino et al. (2019)).

Third, my empirical strategy is original. The idea of applying an RD design on the age of the unemployed, which in turn determines the length of their UI benefits time, is similar to Lalive (2008) and Scrutinio (2019). However, I will be exploiting two age discontinuities instead of one. Furthermore, I will exploit the age discontinuities to check whether hazard rates and survival rates of the control and treatment groups differ. While Card et al. (2007b) used hazard rates differences to check whether two UDs measurement methods differ, we will use them to check whether differences in potential benefits duration bring about UD differences in our control and treatment groups as will be defined in Section 5.

Thus, in light of both the concurrent and conflicting predictions of the selected labor search theory and of the existing empirical literature on the effect of an extension in potential benefits duration on UD and on the path of the empirical re-employment hazard we set out to test the two following hypotheses:

Hypothesis 1 *An increase in potential benefits duration leads to an increase in average unemployment duration.*

Hypothesis 2 *The empirical re-employment hazard follows a U-shaped path during benefits receipt. It then monotonically decreases after benefits exhaustion.*

4 Data Structure, Sampling Strategy, and Descriptive Statistics

4.1 Description of the Data and Questionnaire Structure

The data I will be using for my exploration of the effect of UI benefits duration on UD is called *Rilevazione Continua delle Forze di Lavoro* (RCFL). It is a data set compiled by ISTAT (*The Italian National Statistical Institute*) every trimester Since being introduced at the beginning of the 1950s, the survey has played a primary role in the statistical documentation and analysis of the Italian labor market and has proven to be an indispensable instrument of knowledge for public decision-makers, the media, and citizens alike. The RCFL is harmonized at the European Level according to the REGULATION (EU) 2019/1700 of the European Parliament and of the Council of the European Union, which became operative on the 1st of January 2021. The RCFL is comparable to Labor Force Surveys of other European Countries.

The survey samples around 250 thousand families residing in Italy every year (even if they are temporarily abroad), for a yearly total of about 600 thousand individuals, distributed around approximately 1400 municipalities (the equivalent of US counties). Italian citizens permanently living abroad, or members of communities (i.e. religious communities, military, etc.) are not interviewed. The interviewed families are randomly selected from the *Anagrafe Nazionale della Populazione Residente* (ANPR), i.e. Italian census data. The sampling strategy is a two-stage sampling with stratification of primary sampling units only, aiming at creating a statistically representative sample of the Italian population of interest. The two sampling stages are municipalities and households, respectively. The municipalities are stratified according to their demographic dimension.

The RCFL takes the form of a panel survey conducted every quarter. Families are interviewed 4 times in the period of 15 months. The household rotation scheme is 2-2-2. Families are interviewed consequently for 2 successive quarters, then they are not interviewed for the two following quarters, and finally, they are interviewed again in the 2 final quarters. Hence, the rotation percentage is 50% each quarter. Since transitions from unemployment or retirement to employment for individuals aged 74+ are approximately null, observational units (families) made up only of individuals aged 74+ and unemployed are not re-interviewed.

The RCFL sample survey is continuous insofar as information is collected by interviewers every week of the year, from the 1st of January to the 31st of December of every year. Respondents are interviewed through CAPI (Computer Assisted Personal Interviewing) andor CATI (Computer Assisted Telephone Interviewing) methods. Furthermore, to reduce the chances of losing observational units to follow-up, respondents have a legal obligation to respond to ISTAT's data inquiries according to art. 7 of the *DECRETO LEGISLATIVO 6 settembre 1989, n. 322*, and of the *DECRETO DEL PRESIDENTE DELLA REPUBBLICA 9 marzo 2022*. Nevertheless, since no legal punishment is associated with refusing to answer ISTAT's surveys, we might still see attrition.

The questionnaire is divided into different alphabetically coded sections. If we take the questionnaire used for the 2014 calendar year, we can see that A variables collect information about who answers the survey, namely the informational unit. Such questions entail the personal information of the respondent (question A3), the family member who is responding to the specific survey wave (question A4), and other information about the specific interview setting. B variables collect information about the employment condition of the respondent in the reference week. For instance, question B1 asks whether the respondent has worked at least one hour in the previous work week. B4 questions ask the respondent the type of contract he is under. From section B questions the interviewer determines whether the interviewee is employed or unemployed. If he is in the workforce he moves to section C, which are questions related to the respondent's primary work, and then to section D, which are questions related to the respondent's secondary work. If the interviewee is unemployed, the interviewer asks section E questions, which concern themselves with past work experiences of the respondent. Then, the interviewer moves on to sections F and G questions (respectively work search, and work search agency and services consultation questions) if either the employed or the unemployed respondent can work, while if the unemployed is not able to work he moves directly to section H questions (education and professional training). After the aforementioned bifurcation in the interviewer's questioning pattern, he asks section I, and L questions to every interviewee (respectively UI benefits and other benefits, and family-related questions).

4.2 Sampling Strategy

To capture the effect of an extended potential benefits duration on UD as well as to empirically characterize the shape of re-employment hazard, we exploit two age discontinuities in potential benefits duration concerning the Italian ASpI UI program. In 2014 an unemployed applicant below 50 has a potential benefits duration of 8 months, between 50 and 54 of 12 months, and above 54 of 14 months. In 2015 an unemployed applicant below 50 has a potential benefits duration of 10 months, between 50 and 54 of 12 months, and above 54 of 16 months. Unemployed workers are eligible for a given year's potential benefits duration if their unemployment spell starts in that year[1]. That is a worker whose spell started in 2014, but continues through 2015, will continue to receive benefits according to the 2014 schedule. The goal of our sampling procedure is, then, to restrict our analysis to unemployment spells that started in 2014 or 2015 respectively. To do so we can think of two strategies:

1. Sample agents conditional on them being unemployed (variable *I*1 = 2) at interview wave 1 (variable *W AV QU A = 1*). Thence, follow such agents through the remaining interview waves (*W AV QU A* $\in$ {2, 3, 4}). Establish non-employment duration as the difference between the survey wave at which agents re-gain employment, and the date when they stopped working (variables *E2β* and *E4*).

2. Sample agents who are currently employed (variable *I*1 = 1) and who were previously unemployed (variable *DURRIC* ≠ 999). Then, for each quarterly data-set *RCF* L_l, $l \in \{1, \cdots, L\}$, subset observations to only keep units in their first interview wave (W AV QUA = 1), and so that their current employment duration (variable C21) plus the previous job-search duration (*DURRIC*) is smaller than 3l – 3 and larger than 3*l* – 12.

To carry out Strategy 1 we need to follow the same observational unit through the separate quarterly aggregated *RCF* L_l data sets. Nevertheless, because of privacy legislation, ISTAT does not provide unique identifiers to track down the same agent across quarters, hence making it unfeasible to observe the evolution of unemployment spells in such a manner.

1 The eligibility requirements for ASpI UI benefits are to be an involuntarily unemployed dependent worker (including apprentice workers) with at least 2 years of total contributions to social security and at least 12 months of contributions to social security in the 2 years prior to lay-off. See Section 8 for more details on UI eligibility, the replacement rate, and the change in replacement rate throughout the UI spell.

Another drawback of Strategy 1 is that even if we were able to track the same unit across quarters, we could still end up with right censored spells since the maximum interview period is only 15 months long. Fortunately, for the years 2014 to 2020 only, the variable *W AV QU A* is observable. Hence, Strategy 2 is feasible. It also has the comparative advantage of providing the econometrician with completed job-search durations (variable *DURRIC*), thereby eliminating left or right censoring issues.

The imposed sampling restriction to subset units who are in their first interview wave (*W AV QU A* = 1) only is necessary, otherwise, we would incur in repeatedly sampled observations. The two last sampling conditions make sure that the observations belong to a given calendar year. As an illustrative example, let $l = 1$ such that *RCF* L_1 represents the first quarterly data set of, say, 2014. According to our formula, we would not sample any observations from such a batch of data (*DURRIC* + *C*21 < 0). This is because ISTAT's sampling process is continuous and ISTAT does not publicly circulate the exact date when the interviews take place. Hence, even if we observed a current employment duration (*C*21) plus a previous search duration (*DURRIC*) shorter than the quarter's length (3 months) we could not be sure that it started in 2014. Let, instead, l = 2 such that *RCF* L_2 represents the second quarterly data set of 2014 and *DURRIC* + *C*21 < 3. We easily see that even if agents were interviewed on the first day of the second quarter of 2014 (namely three months and one day away from the 1st of January 2014), then their search duration would still have started in 2014. The last sampling condition (*DURRIC* + *C*21 > 3l – 12) becomes relevant when $l \geq 5$. In the framework of our running example, let's consider $l = 5$. Then *RCF* L_5 represents the first quarter of 2015, with *DURRIC* + *C*21 < 12 and *DURRIC* + *C*21 > 3. Suppose to the contrary that we sampled a unit such that *DURRIC* + *C*21 < 3. If the agents had been interviewed on the last day of the quarter, the search duration would have started in 2015, thereby defeating our aim of sampling search duration started in 2014 only.

The chosen sampling strategy lets us work with completed search durations and lets us identify the year that they started. However, it does not permit to observe whether the agents were receiving UI benefits during their search duration, since the time frame of the questionnaire's only question about benefits receipt (variable *G*9 = 1) is the week preceding the interview itself. Nevertheless, as pointed out in Section 2, the econometrician should use the whole non-employment duration as the outcome variable in order to avoid truncation bias.

4.3 Descriptive Statistics

Now that we have talked about the data structure, we can delve into a hands-on exploration of the data. Table 1 shows the number of observations and the number of variables of each raw RCFL data set that will be used in the analysis. For the years 2014 to 2020, each quarterly data set counts around 90 to 100 thousand observations. The number of reported variables ranges between 340 and 380.

After having carried out our sampling strategy we are left with a data set containing search durations started in 2014, and a data set containing search durations started in 2015. Table 2 shows descriptive statistics for the full samples.

At first glance, we can see that the sampling process severely shrinks the sample sizes. For what concerns the ASpI 2014 data, we are left with a sample size of 324 observations in 7 variables, while for what concerns the ASpI 2015 data, we are left with a sample size of 321 observations in 7 variables. Small sample sizes are the first strong limitation of our data. The outcome variable that will be used in the analysis is the job-search duration in months, while the dependent variable is the worker's age, which is recorded in years. The available covariates are a dummy for gender; a categorical variable for citizenship taking the value of 1 if the unit is an Italian citizen, a value of 2 if the unit is an EU citizen, and taking a value of 3 if the unit is a non-EU citizen; a categorical variable for marital status; a 10-levels categorical variable for educational level; and a categorical variable reporting the geographical region (North, Center, South) of the observed unit.

As we can see from Table 2 age is bounded between 18 and 61, with an average of about 35 for the ASpI 2014 data. On the other hand, age is bounded between 18 and 63, with an average of about 34 for the ASpI 2015 data. The minimum UD in both years is 1 month, while the maximum is 72 in 2014 and 60 in 2015. The average job-search duration is about 15 months in 2014 and about 13.5 in 2015. We see that the standard deviation for the age variable is around 10.5 years in both 2014 and 2015, telling us that the bulk of the job search durations belong to workers aged between around 24.5 and 45.5 in 2014, and 23.5 and 44.5 in 2015. Finally, the standard deviation of job-search duration is about 15 months in 2014 and 14 months in 2015, showing how for both years job-search durations within one standard deviation of the mean fall between 1 month and about 30 months (2.5 years).

Table 1: Outlook of Raw RCFL Quarterly Data

	Year	Quarter	Observations	Number of Variables
1	2014	1	101388	340
2	2014	2	94798	357
3	2014	3	92311	340
4	2014	4	94492	340
5	2015	1	99412	341
6	2015	2	99686	341
7	2015	3	101916	341
8	2015	4	99733	341
9	2016	1	98654	339
10	2016	2	92386	368
11	2016	3	96582	339
12	2016	4	91970	339
13	2017	1	99502	348
14	2017	2	96123	377
15	2017	3	94179	348
16	2017	4	91992	348
17	2018	1	97141	339
18	2018	2	93406	339
19	2018	3	92172	339
20	2018	4	92071	339
21	2019	1	95022	343
22	2019	2	99333	343
23	2019	3	99171	343
24	2019	4	94121	343
25	2020	1	93060	343
26	2020	2	101600	343
27	2020	3	98622	343
28	2020	4	97156	340

Note: Number of observations and of variables of each raw RCFL quarterly data set used for the analysis (period 2014-2020).

Table 2: Summary Statistics ASpI 2014 and 2015

Variable	Mean	St. Dev.	Min	Max
	ASpI 2014 Data			
Sex (SG11 Binary)	1.506	0.501	1	2
Citizenship (CITTAD Categorical)	1.207	0.576	1	3
Age (ETAM Years)	34.716	10.546	18	61
Educational Level (TISTUD Categorical)	5.361	2.340	2	10
Job-search Duration (DURRIC Months)	14.904	15.040	1	72
Marital Status (STACIM Categorical)	1.494	0.789	1	6
Geographical Region (RIP3 Categorical)	1.670	0.790	1	3
Number of Observations				324
	ASpI 2015 Data			
Sex (SG11 Binary)	1.520	0.500	1	2
Citizenship (CITTAD Categorical)	1.237	0.602	1	3
Age (ETAM Years)	34.050	10.546	18	63
Educational Level (TISTUD Categorical)	5.508	2.419	1	10
Job-search Duration (DURRIC Months)	13.502	13.072	1	60
Marital Status (STACIM Categorical)	1.439	0.625	1	3
Geographical Region (RIP3 Categorical)	1.726	0.844	1	3
Number of Observations				321

Note: Descriptive statistics for the full samples of ASpI job-search durations started in 2014 and 2015 (own calculations based on RCFL publicly available data).

Now that we have analyzed the full samples, we can gain more information about the data by looking at summary statistics conditional on age. More specifically, we divide the full samples into the three age groups that correspond to the different potential benefits durations (49 years or younger, between 50 and 54, and 55 years or older) and calculate the mean, median, and standard deviation of job-search duration. As we can see from Table 3 the number of available observations are unevenly split between age groups both in 2014 and 2015. In 2014 the first age group, unemployed workers younger than 50, counts 284 observations out of the total of 324, the second, unemployed workers between 50 and 54, counts 30 observations, and the last, unemployed workers older than 54, counts a mere 10 observations. In 2015 the first age group counts 290 observations, the second counts 21 observations, and the last counts 10 observations. This means that around 90% of all job search durations belong to the first age group in both years. The fact that the treatment groups (50 to 54, and 55 or older) represent about 10% of our data is the second serious limitation of the data. The culprit of such a stark decrease in the number of observations for what concerns the two latter age groups is most probably the early retirement age of the Italian pension system. The earliest a worker could have retired in Italy in 2014 and 2015 was 62 OECD (2015), making the longer potential benefits durations at age 55 and older a bridge to retirement. For a detailed study on the relationship between UI benefits and early retirement see Tatsiramos (2007); Lalive (2008); Inderbitzin et al. (2016). For the purpose of our paper, we will be contempt to acknowledge that the estimates of the effect of an extension in potential benefits duration on UD will have to be taken extremely cautiously when comparing the 50-54 and 55 and older age groups.

Table 3: Summary Statistics of ASpI 2014 and 2015 by Age Groups

	Age Groups		
	ASpI 2014 Data		
Job-search Duration (DURRIC Months)	< 50	50-54	> 54
Mean	13.45	22	43.8
Median	10	22	30
Standard Deviation	13.96	17.18	19.14
Number of Observations	284	30	10
	ASpI 2015 Data		
Mean	12.81	19.19	21.7
Median	11.5	20	12.5
Standard Deviation	12.62	14.97	17.16
Number of Observations	290	21	10

Note: Descriptive statistics for job-search durations conditional on age (younger than 50, between 50 and 54, older than 54) for the years 2014 and 2015 (own calculations based on RCFL publicly available data).

Said that Table 3 shows how the average job-search duration increases in the age groups for both 2014 and 2015. This is the first rough indication that we might expect to find positive estimates for the effect of potential benefits duration on UD. Finally, while it is true that the median length of job-search durations conditional on the age groups increases for the year 2014, this is not so for the year 2015, suggesting that the higher average as compared to the other age groups could be due to the presence of outliers.

5 Econometric Framework

Our goal is to ascertain the effect of an extension in potential benefits duration on the length of unemployment duration, as well as to characterize the empirical path of the employment hazard. In this section, we develop econometric methodologies that will allow us to analyze such effects.

5.1 Econometric Model for the Examination of Hypothesis 1

We can consider the two age discontinuities provided by the Italian insurance system ASpI in the context of the potential outcomes Rubin Causal Model (RCM), and use them to estimate the Local Average Treatment Effect (LATE) of an extended benefit duration, following Imbens and Lemieux (2007). Let $Y_i(0)$ and $Y_i(1)$ be the potential outcomes for agent *i*, whereby 0 corresponds to non-treatment (in our context being eligible for a lower UI benefit duration), and 1 to treatment (in our context being eligible for a higher UI benefit duration). If we denote the indicator $W_i \in (0, 1)$ as taking the value of 0 if unit *i* was not treated, and the value of 1 if unit i was treated, then the observed outcome is

$$Y_i = (1 - W_i) \cdot Y_i(0) + W_i \cdot Y_i(1) = \begin{cases} Y_i(0) & \text{if } W_i = 0 \\ Y_i(1) & \text{if } W_i = 1 \end{cases}$$

Furthermore, we observe the co-variate X_i, our forcing variable age and an M-vector of controls Z_i. Hence for each individual in our investigation, we observe the quadruple (Y_i, W_i, X_i, Z_i) obtained from our samples. Since the assignment of the treatment ($W_i = 1$), namely having a longer UI benefits duration, is determined by age (X_i) we use a Sharp Regression Discontinuity (SDR) design. Here we assume that the treatment W_i, namely having a longer potential benefits duration, is a function of the covariate age X_i, for individual $i \in I$, for $I = 1 \ldots N$. Formally, we have

$$W'_{ij} = \begin{cases} 0 & \text{if } X_{ij} < C_1 \\ 1 & \text{if } C_1 \leq X_{ij} < C_2 \\ 2 & \text{if } X_{ij} \geq C_2 \end{cases}$$

where $C_1 = 50$, $C_2 = 55$, for $W_{ij} = (0, 1, 2)$ meaning that the agent had a UI benefits duration of 8, 12, and 14 months respectively in year $j = 2014$, and for $W_{ij} = (0, 1, 2)$ meaning that the agent had a UI benefits duration of 10, 12, and 16 months respectively in year $j = 2015$. In practice, in our estimation, we will use the control group defined by $X_{ij} < C_1$, when comparing it to the treatment $C_1 \leq X_{ij} < C_2$, but we will use the control group $C_1 \leq X_{ij} < C_2$ when comparing it to the treatment group $X_{ij} \geq C_2$.

Imbens and Lemieux (2007) show that assuming continuity of conditional regression functions, namely that $\mathbb{E}[Y_i(0) \mid X]$, and $\mathbb{E}[Y_i(1) \mid X]$ are continuous in *X*, and trivially assuming $Y_i(0), Y_i(1) \perp W_i \mid X_i$, namely that the potential outcomes are independent of treatment conditional on the forcing variable, the local average treatment effect satisfies $\tau_{\text{SRD}} = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = c] = \lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x]$.

We can obtain the LATE of being in the longer potential benefits duration group by estimating a local linear (LL) regression schedule. We estimate a LL regression model with a triangular kernel and with smoothing parameter *h*. Fan and Gijbels (1994); Porter (2003) showed that using a triangular kernel is convergence rate optimal, and such a method is widely used in the literature. Furthermore, we use Imbens and Kalyanaraman (2012) MSE (Mean Squared Error or Risk) minimizing formula to find our optimal bandwidth *h*. Our specification is

$$\min_{\alpha,\beta,\tau,\gamma} \sum_{i=1}^{N} 1\{c - h \leq X_{ijk} \leq c + h\} \cdot (Y_{ijk} - \alpha - \beta \cdot (X_{ijk} - c) - \tau \cdot W_{ijk} - \gamma \cdot (X_{ijk} - c) \cdot W_{ijk})^2$$

where Y_{ijk} is the completed UD spell of individual *i* filing for UI benefits in year *j* considering the discontinuity $k \in (C_1, C_2)$, X_{ijk} is the age of agent *i* filing for UI benefits in year *j* considering the discontinuity $k \in (C_1, C_2)$, W_{ijk} indicates whether the individual is in the treatment or control groups as defined above. In order to use our estimating equation for years $j \in (2014, 2015)$, we will treat the two discontinuities in age each year separately, thereby using it four times. In the context of our hypotheses, finding a positive value on the τ coefficient would indicate that an extension in potential benefits duration has a positive causal effect on UD, thereby validating Hypothesis 1.

5.2 Econometric Model for the Examination of Hypothesis 2

On top of estimating the LATE of a 1-week increase in UI benefits, we delineate the path of the empirical hazard and control for spikes at benefits exhaustion. We identify our control and treatment groups (g_{1j}, g_{2j}, g_{3j}) for $j \in$ (2014, 2015) such that the age of g_{1j} is between 45 and 49, the age of g_{2j} is between 50 and 54, and the age of g_{3j} is between 55 and 60. We assume that these groups are close enough in age to be appropriate control and treatment groups. We then estimate the hazard and the survival of $g_{1j} - g_{2j}$, and of $g_{1j} - g_{2j}$. If Meyer (1990); Katz and Meyer (1990a,b); Moffitt (1985) finding of a spike at benefits exhaustion holds, we expect to see a change in the hazard differences when the benefits of the first group expire before the benefits of the second groups. Furthermore, we test whether the hazard rate stays constant after benefits expiration as in Mortensen (1977) or decreases as suggested by DellaVigna et al. (2017).

We estimate individuals' hazard and survival functions using Kaplan-Meyer methods because they are robust to right censoring and we are not interested in hazard and survival analysis far from benefits exhaustion. After Singer and Willett (1993); Efron (1988), we consider a setup where we denote by T the set of all discrete unemployment durations, and $t_i \in \mathbb{T}$, where $\mathbb{T} = (t_1, \ldots, t_T)$ is the ordered support of discrete times when an unemployment duration could end. Since unemployed people could leave unemployment forever, here we have $t_T = +\infty$. We also have a non-censoring indicator $D_i = 1(T_i \leq S)$, where $S \in \mathbb{T}$ represents either the time of the last RCFL wave or the time when agents are lost to follow up because of non-response. We have $D = 0$ if $T_i > S$, and $D = 1$ if $T_i \leq S$. Because of censoring, we do not observe T_i if $T_i > S$. What we do observe is a variable $H_i = min(T_i, S_i)$. Furthermore, we assume that $T_i \perp S_i$, namely that censoring is uninformative, i.e. it does not co-vary with unemployment duration. Our assumption seems valid since we do not expect a covariance channel between an individual's interview period ending and their likelihood of finding a new job. With this information, we can estimate the true survival function $S(t_i) = Pr(T > t_i)$ using the Kaplan-Meyer estimator, namely

$$\hat{S}(t_i) = \prod_{k=1}^{i}(1 - \hat{\lambda}(t_i))$$

and the true hazard function $\lambda(t_i) = P(T_i = t_i | T_i \geq t_i)$ by

$$\hat{\lambda}(t_i) = \frac{\mathrm{P}(T_i \geq t_i,\ T_i \geq h_i)}{\mathrm{P}(H_i \geq t_i)} = \frac{\mathbb{E}[D \cdot \mathbf{1}(H = t)]}{\mathbb{E}[\mathbf{1}(H \geq t)]}$$

Due to the small sample sizes we only engage in a characterization of the

empirical hazard and survival functions, without doing inference. Hence, our evidence in support or against Hypothesis 2 will be solely descriptive.

6 Empirical Results

We now delve into the results obtained from our analysis.

6.1 Empirical Evidence in Support of Hypothesis 1

Tables 4 and 5 show the regression results of our main specification. Table 4 comprises the estimates of the SRD design on the discontinuities in potential benefits duration at ages 49 and 54 for the ASpI 2014 data. Table 5 exhibits equivalent calculations for the ASpI 2015 data. Tables 4 and 5 report both the total and the week-per-week effects of an 'extension' in potential benefits duration.

Table 4: Sharp Regression Discontinuity Design - ASpI 2014 Data

	ASpI 2014 Data	
	Discontinuity at Age 49	
	Total Effect (Months)	Week-per-Week
LATE (DURRIC)	18.61**	1.07**
	(8.364)	(0.671)
Bandwidth	7.157	
Observations	70	
F Statistic	3.659**	
	(df = 3; 66)	
	Discontinuity at Age 54	
	Total Effect (Months)	Week-per-Week
LATE (DURRIC)	43.23***	4.974***
	(15.87)	(1.826)
Bandwidth	2.548	
Observations	18	
F Statistic	4.598**	
	(df = 3; 14)	

*p<0.1; **p<0.05; ***p<0.01

Note: Results of the SRD design using Imbens and Kalyanaraman (2012) formula for bandwidth selection. Regressing job-search durations (DURRIC) on age (ETAM) for the discontinuities in the ASpI 2014 program at ages 49 and 54. Using a triangular kernel and heteroskedasticity-consistent standard errors. Reporting both the full effect and the week-per-week effect of an increase in potential benefits duration. The week-per-week effect is calculated by dividing the full effect by the 'extension' in potential benefits duration corresponding to the age discontinuity and by the number of weeks in a month (4.34524). Standard errors in parentheses.

Table 5: Sharp Regression Discontinuity - ASpI 2015 Data

	ASpI 2015 Data	
	Discontinuity at Age 49	
	Total Effect (Months)	Week-per-Week
LATE (DURRIC)	-10.115 (9.008)	-1.164 (1.036)
Bandwidth	5.432	
Observations	55	
F Statistic	1.689 (df = 3; 51)	
	Discontinuity at Age 54	
	Total Effect (Months)	Week-per-Week
LATE (DURRIC)	0.8506 (12.78)	0.049 (0.735)
Bandwidth	5.305	
Observations	30	
F Statistic	0.07901 (df = 3; 26)	
Note:		*p<0.1; **p<0.05; ***p<0.01

Note: Results of the SRD design using Imbens and Kalyanaraman (2012) formula for bandwidth selection. Regressing job-search durations (DURRIC) on age (ETAM) for the discontinuities in the ASpI 2015 program at ages 49 and 54. Using a triangular kernel and heteroskedasticity-consistent standard errors. Reporting both the full effect and the week-per-week effect of an increase in potential benefits duration. The

week-per-week effect is calculated by dividing the full effect by the 'extension' in potential benefits duration corresponding to the age discontinuity and by the number of weeks in a month (4.34524). Standard errors in parentheses.

As we can see, the sign of the LATE is positive for all of the age discontinuities each year, except for the discontinuity at age 49 in 2015. Furthermore, the coefficients for 2014 are both statistically significant (the one pertaining to the discontinuity at age 49 at the 0.05 significance level, and the one pertaining to the discontinuity at age 54 at the 0.01 significance level). Obtaining such a result for three of the discontinuities is surprising given the limited nature of our data. As argued in Section 5 the estimates in Tables 4 and 5 represent the LATE of having been treated with an extended benefits duration. If we look at the upper quadrant of Table 4, we see that the total average effect of four additional months of potential benefits duration on UD is 18.61 months or about 1.5 years (remember workers below age 50 were entitled to 8 months of UI benefits, and those between ages 50 and 54 were entitled to 12 months of UI benefits in 2014). This means that, on average, the increase in UD of a 1-week increase in potential benefits duration is about 1.07 weeks. The LATE estimate in the lower quadrant of Table 4 tells us that on average the effect of a two months extension in potential benefits duration on UD is 43.23 months, or about 3.6 years (workers above age 54 were entitled to 14 months of UI benefits). This means that, on average, the increase in UD of a 1-week increase in potential benefits duration is about 4.97 weeks. In the upper quadrant of Table 5, which corresponds to the discontinuity at age 49 in 2015, the LATE estimate tells us that on average the effect of a two months extension in potential benefits duration is a decrease of UD of 10.115 months (workers below age 50 were entitled to 10 months of UI benefits, and those between ages 50 and 54 were entitled to 12 months of UI benefits in 2015). This means that, on average, the decrease in UD of a 1-week increase in potential benefits duration is about 1.164 weeks. In the lower quadrant of Table 8, which corresponds to the discontinuity at age 54 in the year 2015, the LATE estimate tells us that on average the effect of a four months extension in potential benefits duration is an increase of UD of about 1 month (workers above age 54 were entitled to 16 months of UI benefits). This means that, on average, the increase in UD of a 1-week extension in potential benefits duration is about 0.049 weeks.

A *prima facie* consideration of our results seems to be validating the positive relationship between extended potential benefits duration and increased UD for the year 2014, but we obtain a negative estimate for the discontinuity at age 49 in the year 2015. Furthermore, it seems that for the most part, the magnitudes of our results match neither the pioneering nor the more recent studies, neither in the US nor in Europe. In fact, as we have just seen, we have obtained week-per-week effects of about 1.07 and 4.97 weeks for the year 2014, 0.049 weeks for the second discontinuity in 2015, and a negative estimate for the first discontinuity in 2015. We can think of two explanations for this mismatch. The first is the quality of the data, the second is, again, the fact that estimates using the discontinuities at age 54 could be inflated due to the role of UI benefits as an avenue to early retirement. For what concerns the first explanation, we have already noted in 4 that the sampling procedure severely shrinks our data. Furthermore, while the estimates using the discontinuity at age 54 could already suffer from an upward bias due to the early retirement hypothesis, the number of observations used for their calculation is extremely low (18 and 30 data points). Even if we cannot unfortunately ameliorate the situation of our sample sizes, we can work the opposite way by repeating our analysis by removing clear job-search outliers. Upon an examination of the upper tail of the distribution of job-search durations, we notice that about 3% of the data consists of durations 48 months or longer (4 years) both in 2014 and 2015. As a comparison point, the longest potential benefits duration in the studied period was merely 16 months. Hence, with the suspicion that such outliers could severely affect the LATE estimates in our small sample size environment and with the plausible assumption that such disproportionately long durations are most probably not as sensitive to changes in potential benefits durations, we remove them. We then re-run our SRD calculations for both 2014 and 2015.

Table 6: Sharp Regression Discontinuity Design - ASpI 2014 Data (No Outliers)

	ASpI 2014 Data	
	Discontinuity at Age 49	
	Total Effect (Months)	Week-per-Week
LATE (DURRIC)	11.41	0.656
	(8.170)	(0.470)
Bandwidth	6.229	
Observations	54	
F Statistic	1.638	
	(df = 3; 50)	
	Discontinuity at Age 54	
	Total Effect (Months)	Week-per-Week
LATE (DURRIC)	30.63**	3.524**
	(13.784)	(1.586)
Bandwidth	2.221	
Observations	17	
F Statistic	2.884	
	(df = 3; 13)	

*p<0.1; **p<0.05; ***p<0.01

Note: Results of the SRD design using Imbens and Kalyanaraman (2012) formula for bandwidth selection. Regressing job-search durations (DURRIC) on age (ETAM) for the discontinuities in the ASpI 2014 program at ages 49 and 54. Using a triangular kernel and heteroskedasticity-consistent standard errors. Reporting both the full effect and the week-per-week effect of an increase in potential benefits duration. The week-per-week effect is calculated by dividing the full effect by the 'extension' in potential benefits duration corresponding to the age discontinuity and by the number of weeks in a month (4.34524). Removing observations with an observed job-search duration longer than 48 months (4 years). Effectively removing 2.78% of the data. Standard errors in parentheses.

Table 7: Sharp Regression Discontinuity - ASpI 2015 Data (No Outliers)

	ASpI 2015 Data	
	Discontinuity at Age 49	
	Total Effect (Months)	Week-per-Week
LATE (DURRIC)	0.1455	0.017
	(6.362)	(0.728)
Bandwidth	8.992	
Observations	75	
F Statistic	1.7864	
	(df = 3; 71)	
	Discontinuity at Age 54	
	Total Effect (Months)	Week-per-Week
LATE (DURRIC)	15.30**	0.88**
	(7.686)	(0.442)
Bandwidth	4.561	
Observations	23	
F Statistic	0.7987	
	(df = 3; 19)	
Note:		*p<0.1; **p<0.05; ***p<0.01

Note: Results of the SRD design using Imbens and Kalyanaraman (2012) formula for bandwidth selection. Regressing job-search durations (DURRIC) on age (ETAM) for the discontinuities in the ASpI 2014 program at ages 49 and 54. Using a triangular kernel and heteroskedasticity-consistent standard errors. Reporting both the full effect and the week-per-week effect of an increase in potential benefits duration. The week-per-week effect is calculated by dividing the full effect by the 'extension' in potential benefits duration corresponding to the age discontinuity and by the number of weeks in a month (4.34524). Removing observations with an observed job-search duration longer than 48 months (4 years). Effectively removing 1.87% of the data. Standard errors in parentheses.

Tables 6 and 7 show the result of our exercise. We can see that the coefficient of the 2015 discontinuity at age 49 turns positive, that the coefficient of the 2014 discontinuity at age 49 looses statistical significance, and that the coefficient of the 2015 discontinuity at age 54 becomes statistically significant at the 0.05 level. Apart from the coefficient of the 2015 discontinuity at age 49, the magnitude of the other coefficients considerably decreases. This attests to our sense that the full sample estimates could have been dragged up by outliers.

Table 6 shows the 2014 LATE estimates, obtaining week-per-week magnitudes of 0.656 and 3.524 for the discontinuities at age 49 and 54 respectively. Table 7 shows the 2015 LATE estimates, obtaining week-per-week magnitudes of 0.017 and 0.88 for the discontinuities at age 49 and 54 respectively. Hence, our estimates of the causal effect of an extension in potential benefits duration on UD range between 0.02 and 3.5. Nevertheless, we argued that we need to be cautious about the estimates obtained utilizing the discontinuity at age 54. Hence, if we only focus on the estimates related to the discontinuity at age 49 and use them as a lower and upper bound for our effect, we are left with a range bounded by 0.02 and 0.66.

It is interesting to notice that, albeit significantly larger, such a range is fairly close to the one calculated by Citino et al. (2019) in the context of the Italian UI program in the years 2009-2012 (0.12-0.55). Of course, here we are constructing our interval using the estimates only. If we look at the standard errors for our estimates we can see that 95% confidence intervals would be even larger.

Figure 1 shows 95% confidence intervals for our estimates without outliers. As we can see confidence intervals are large for all estimates. They are larger for estimates that exploit the discontinuity at age 54 for both years. Furthermore, confidence intervals for estimates exploiting the discontinuity at age 49 for both years include 0. We conclude that, while it makes sense to only consider the estimates exploiting the discontinuity at age 49 due to possible overestimation of the effect for estimates exploiting the discontinuity at age 54, our causal design does not allow us to rule out the possibility of a non-finding

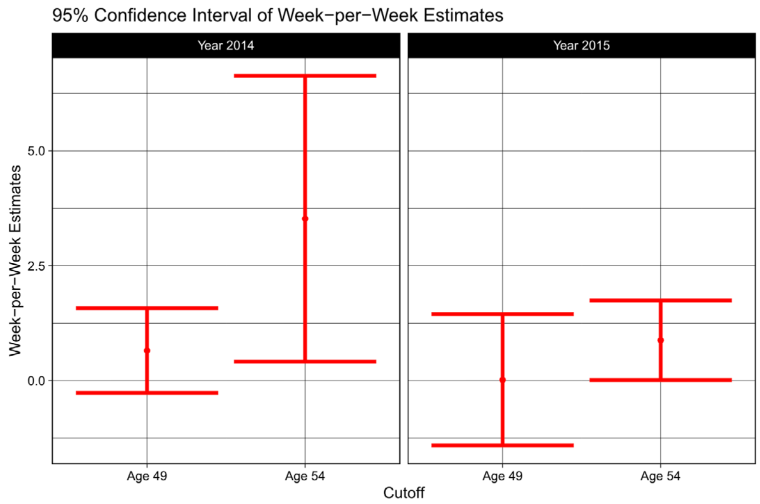


Figure 1: Plot of the SRD week-per-week estimates with 95% confidence intervals. Using ASpI 2014 and 2015 data without outliers (filtering out search durations longer than 48 months).

6.2 Evidence in Support of Hypothesis II

We can now focus on the analysis of Figures 2 to 8. Figures 2 and 3 portray the hazard functions of groups (g_{1j}, g_{2j}, g_{3j}) for as defined above for years $j \in \{2014, 2015\}$ respectively. Figures 4 and 5 portray the survival functions. Figures 6 and 7 portray the difference in hazard rates of groups $g_{1j} - g_{2j}$ and $g_{2j} - g_{3j}$ for $j = 2014$. Figures 8 and 9 portray the difference in hazard rates of groups $g_{1j} - g_{2j}$ and $g_{2j} - g_{3j}$ for $j = 2014$. We have omitted the confidence intervals from Figures 2 to 5 for the sake of visual clarity. Nevertheless, variance and 95% CIs for both the hazard and the survival functions can be found in Tables 11 to 16 in Section 8. Because of data limitations, there are several months whereby no UDs end (in 2014 there are 8 different duration lengths for what concerns g_1 and g_2, and 5 for what concerns g_3; in 2015 there are 10 different duration lengths for what concerns g_1, 6 for what concerns g_2, and 4 for what concerns g_3). For the sake of our analysis, we are assuming that the hazard rate at a given month with a missing UD is the same as the hazard at the first month that precedes it without missing observations (even if formally the hazard at these months would be 0 since we estimate it by dividing the number of people who find a job by the number of people who are still in the pool of

the unemployed at every month).

If we look at Figures 2, we do not seem to find evidence of spikes at benefit exhaustion for what concerns Groups 1 and 3. The hazard rate of Group 1 around month 8 seems to be flat, as well as the hazard rate of Group 3 around month 14. Nonetheless, the hazard rate of Group 2 at month 12 sees a significant increase of about 0.1. Similar results hold for Figure 3, there does not seem to be a significant uptake in the escape rate before months 10, and 14, but the hazard of Group 2 increases significantly around month 12. These findings are tentative and descriptive due to the nature of our data. Having more granular micro-data allowing us to track the hazard of unemployed workers month by month, or even day by day would make our estimation more precise.

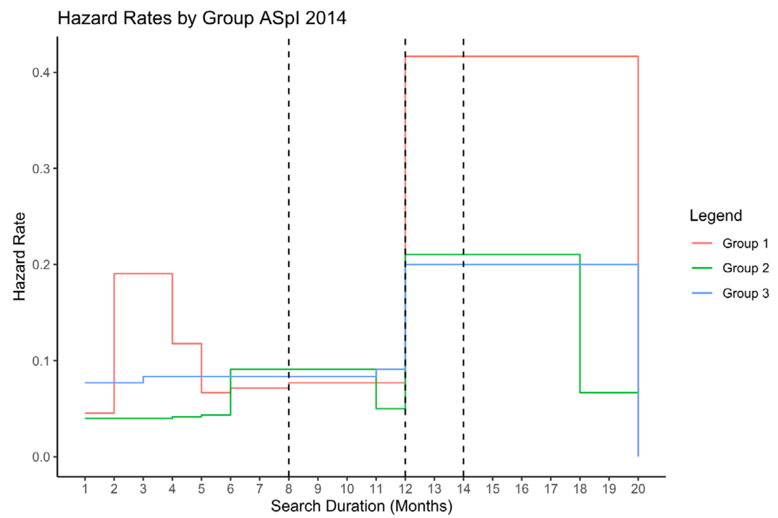


Figure 2: Plot of re-employment hazard functions for workers aged between 45 and 49 (Group 1), between 50 and 54 (Group 2), and aged between 55 and 60 (Group 3) for the year 2014.

A more surprising result is seeing that the hazard and survival functions of the three groups in Figures 4 and 5 differ significantly even before the 8th month in the case of 2014 and before the 10th month in the case of 2015. In Figure 4 we see that the survival function of Group 1 is significantly lower than the survival functions of both Groups 2 and 3 throughout the whole 20 months. In Figure 5, instead, such difference is not as demarcated. The three survival functions intersect each other multiple times, but the one of Group 1 ultimately ends up at the bottom after month 12. Figure 4 suggests that workers might anticipate a longer benefits duration, thereby decreasing their search effort during the initial phase of their unemployment duration as in Card et al. (2007a). The latter finds that in Austria job seekers who are eligible for a longer benefits duration exhibit job finding rates during the initial phase of their unemployment duration 5%–9% lower than those who are eligible for the shorter benefits duration.

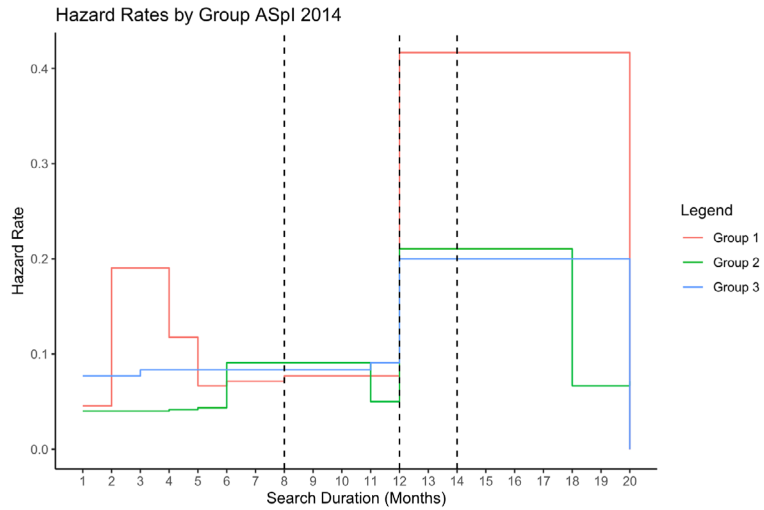


Figure 3: Plot of re-employment hazard functions for workers aged between 45 and 49 (Group 1), between 50 and 54 (Group 2), and aged between 55 and 60 (Group 3) for the year 2015.

If we now look at figures 6 to 9, we see some conflicting results. Figure 6 shows the difference between groups $g_{1j} - g_{2j}$ for $j = 2014$. We see that up to month 4, there is a positive hazard difference in the two groups, but then between months 4 and 12 is almost zero. It then shoots up to 0.2 up to month 17. We have expected an increase in this difference around month 8, since if workers actually exit unemployment *en masse* when they are closer to benefit exhaustion, then the hazard of workers in group 1 would increase, while the hazard of workers in group 2 would remain constant, thereby increasing the difference in hazards. Nevertheless, we do not seem to see this increase in month 8. We do see an increase in the difference in hazards slightly before and right after month 12. This increase goes against our hypothesis that hazard should decrease after benefits exhaustion as found by Card et al. (2007b); DellaVigna et al. (2017).

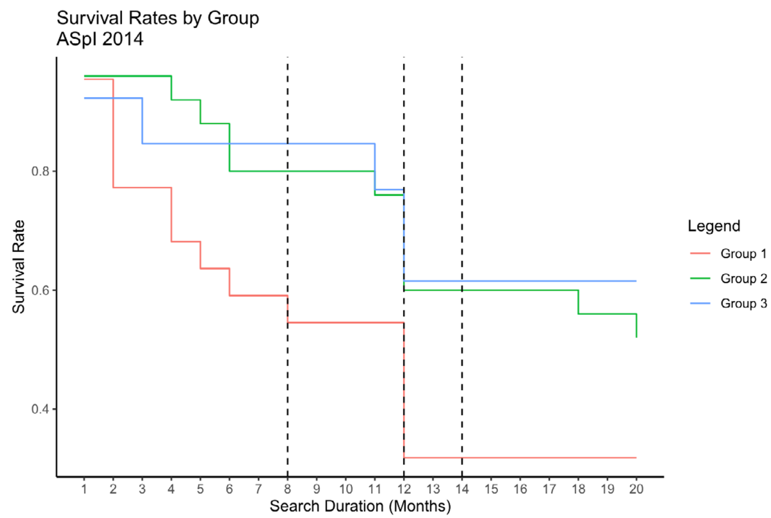


Figure 4: Plot of re-employment survival functions for workers aged between 45 and 49 (Group 1), between 50 and 54 (Group 2), and aged between 55 and 60 (Group 3) for the year 2014.

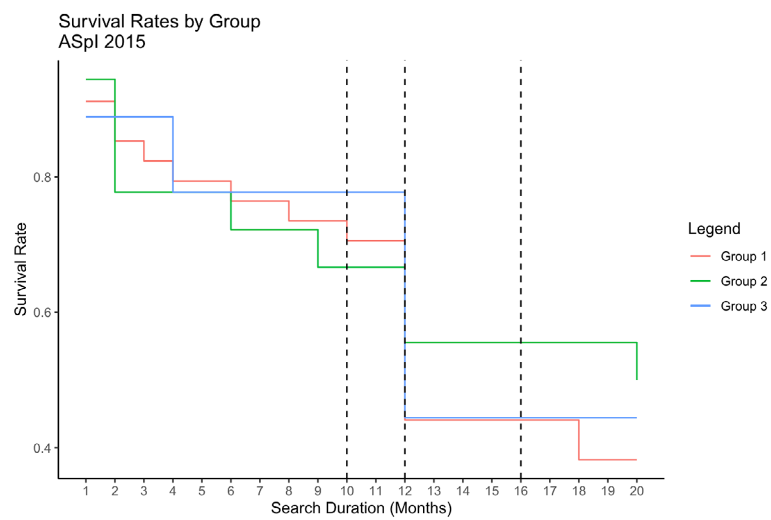


Figure 5: Plot of re-employment survival functions for workers aged between 45 and 49 (Group 1), between 50 and 54 (Group 2), and aged between 55 and 60 (Group 3) for the year 2014.

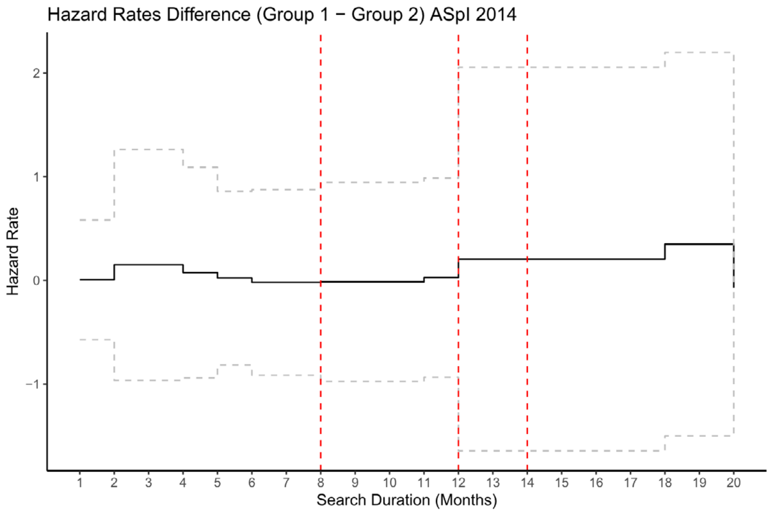


Figure 6: Plot of the difference in re-employment hazard between workers aged between 45 and 49 (Group 1) and workers aged between 50 and 54 (Group 2) for the year 2014. 95% confidence interval calculated using Greenwood (1927) formula.

In Figure 7 we see that the difference in hazards between groups 2 and 3 is close to 0 up until month 18. We see a slight decrease in hazard before month 12, which is counterintuitive, since according to our predictions, the hazard of group 2 should increase before benefit exhaustion, while the hazard of group 3 should remain constant. We do not see a significant difference between the two groups before month 14 either.

Finally, again, Figure 8 does not seem to confirm the hypotheses present in the literature. We see a close to 0 difference between the hazards, and then a significant increase in the hazard differences after month 12. Figure 9 is also puzzling because we see a very strong negative difference after month 12 and no signs of spikes before months 12 and 16.

We conclude that the empirical hazard functions and hazard differences do not seem to provide evidence for Hypothesis 2. However, we have already mentioned that our evidence in support of Hypothesis 2 would be descriptive and severely limited by the data at hand.

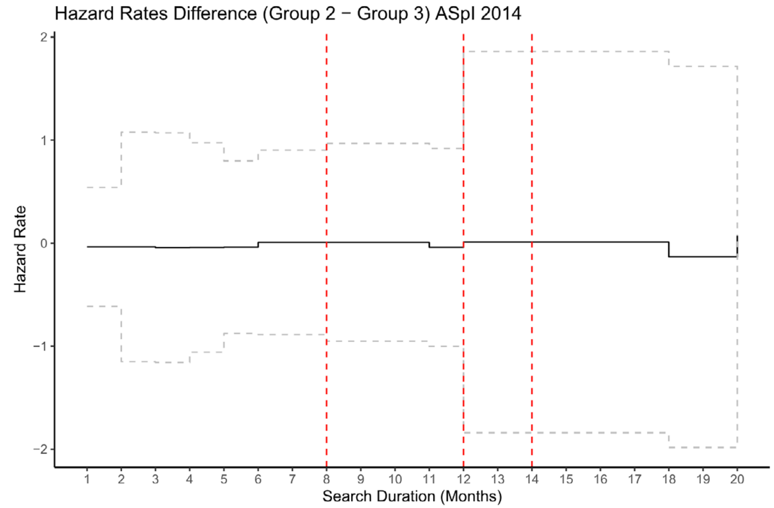


Figure 7: Plot of the difference in re-employment hazard between workers aged between 50 and 54 (Group 2) and workers aged between 55 and 59 (Group 3) for the year 2014. 95% confidence interval calculated using Greenwood (1927) formula.

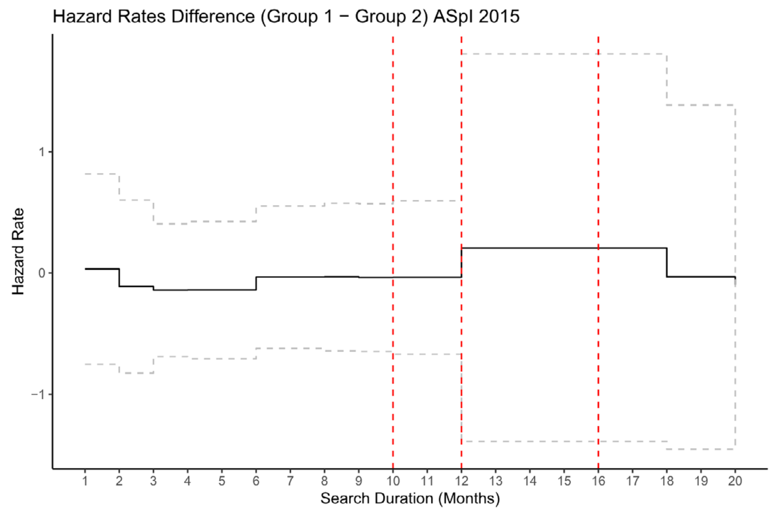


Figure 8: Plot of the difference in re-employment hazard between workers aged between 45 and 49 (Group 1) and workers aged between 50 and 54 (Group 2) for the year 2015. 95% confidence interval calculated using Greenwood (1927) formula.

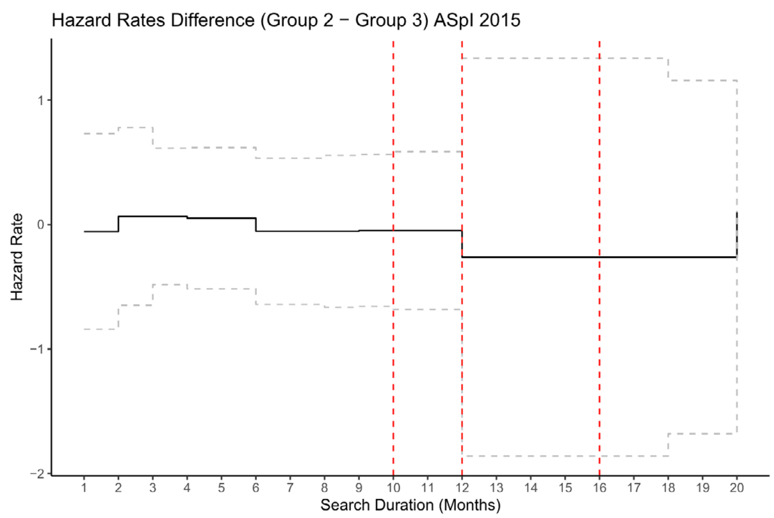


Figure 9: Plot of the difference in re-employment hazard between workers aged between 50 and 54 (Group 2) and workers aged between 55 and 59 (Group 3) for the year 2014. 95% confidence interval calculated using Greenwood (1927) formula.

7 Robustness Checks and Sensitivity Analysis

We now check the robustness and sensitivity of our results. First of all, as Imbens and Lemieux (2007); Lee and Lemieux (2010) point out and as we have shown in Section 5, an SRD can be effectively thought of as a randomization experiment for units close enough to the threshold of the forcing variable. Hence, we can check for the validity of such a randomization experiment by first looking at the continuity of the forcing variable around the threshold, and second by looking at whether baseline covariates are continuous around the threshold. This makes sense because an SRD assumes that the probability of the randomization assignment is 0 below the threshold of the forcing variable and 1 above the forcing variable, hence if the density of the forcing variable were to be discontinuous around the threshold, we could not think in a *ceteris paribus* framework, whereby the only difference between the two groups is given by their age at the threshold.

7.1 Checking for Continuity of the Forcing Variable

For what concerns the first check, we can rely on both visual evidence and formal continuity density tests relying on software created by Cattaneo et al. (2018, 2022). Figures 10 to 13 show density histograms of the values of the forcing variable around the thresholds (age 49 and 54), as well as the estimated density using a local-polynomial density estimator based on the CDF of the observed sample for both 2014 and 2015. As we can see in figures 10 and 11 the estimated density seems to be fairly close below and above the thresholds.

Figures 12 and 13 show the two estimated discontinuities are further apart. Neverthe less, while the CIs of the estimated density below and above the threshold are large, they significantly overlap.

While visual evidence is useful, ultimately we need to rely on formal evidence of continu ity. We look at a formal density test attributed to McCrary (2008), computed using software by Cattaneo et al. (2018, 2022). In such a test LL smoothing of density histograms is used.

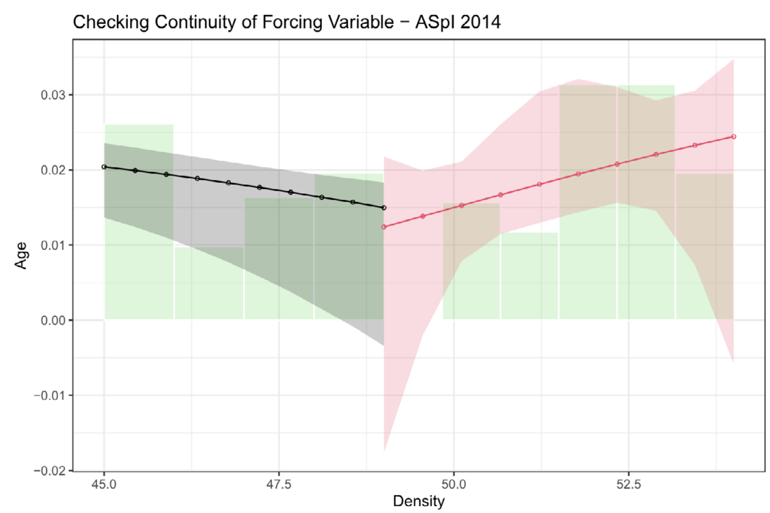


Figure 10: McCrary (2008) style density test performed using software by Cattaneo et al. (2018, 2022) for the discontinuity at age 49 in the ASpI 2014 UI program.

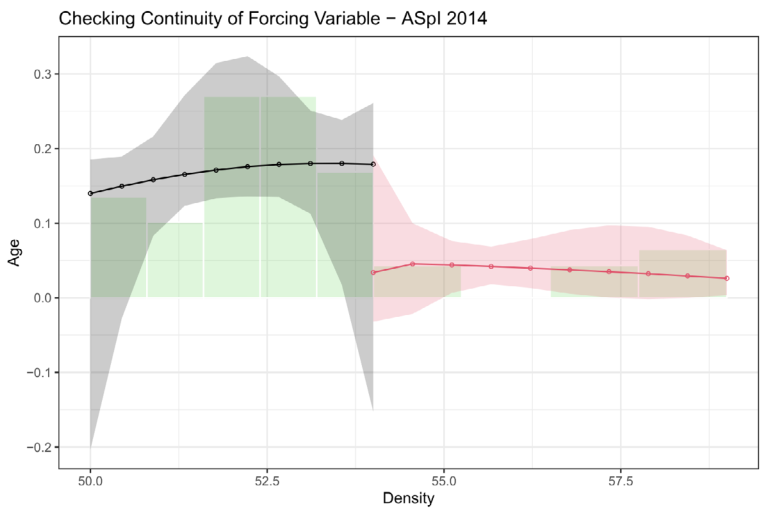


Figure 11: McCrary (2008) style density test performed using software by Cattaneo et al. (2018, 2022) for the discontinuity at age 54 in the ASpI 2014 UI program.

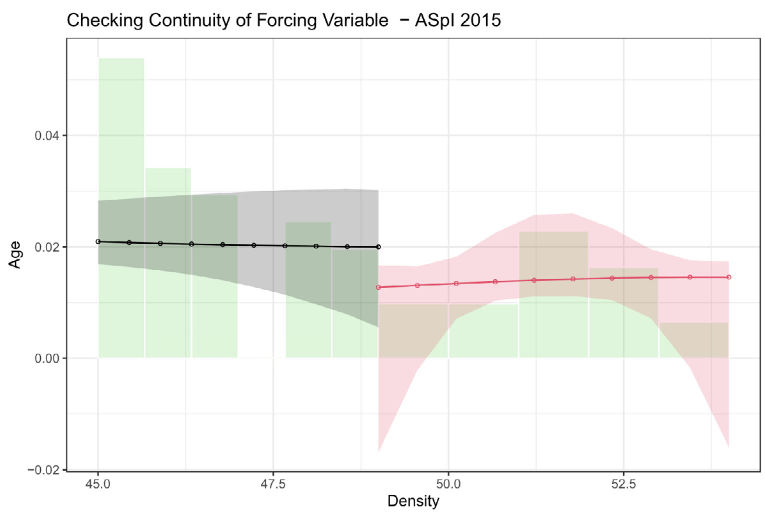


Figure 12: McCrary (2008) style density test performed using software by Cattaneo et al. (2018, 2022) for the discontinuity at age 49 in the ASpI 2015 UI program.

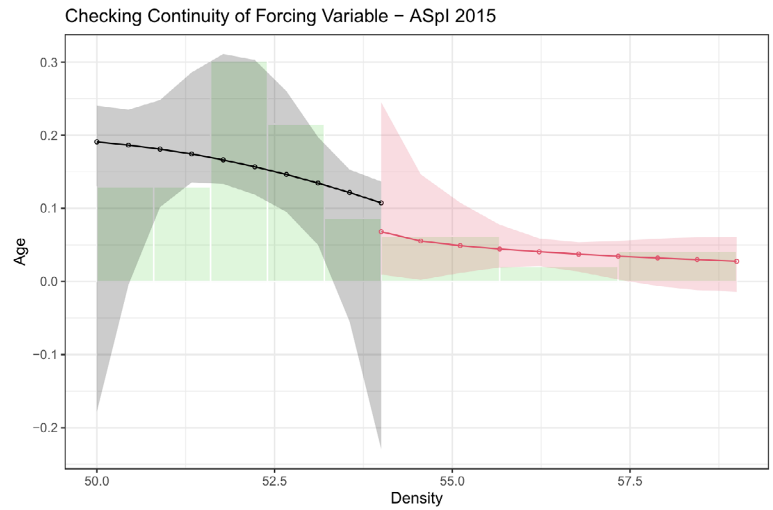


Figure 13: McCrary (2008) style density test performed using software by Cattaneo et al. (2018, 2022) for the discontinuity at age 54 in the ASpI 2015 UI program.

The midpoints of the histogram bins are treated as a regressor, and the normalized counts of the number of observations falling into the bins are treated as an outcome variable. To accommodate the potential discontinuity in the density, local linear smoothing is conducted separately for the bins to the right and left of the point of potential discontinuity. If we look at Table 8 we see the result of such a density test.

Table 8: Formal McCrary Density Test

Age Discontinuity	T-Value	P-Value
ASpI 2014 Data		
49	-0.2236	0.8231
54	0.2718	0.7858
ASpI 2015 Data		
49	-1.839	0.0659
54	1.0002	0.3172

Note: Results of a formal McCrary (2008) style density test performed using software developed by software by Cattaneo et al. (2018, 2022) for discontinuities at age 49 and 54 in both 2014 and 2015.

We notice that for all four discontinuities under consideration (ages 49 and 54, years 2014 and 2015) the P value of the test is significantly larger than 0.05, which tells us that we do not have a significant difference between the estimated densities below and above the thresholds. It must be noted, nevertheless, that the P values for the 2015 discontinuities are significantly lower than the 2014 ones. While our formal test is reassuring, evidence tells us that even if our forcing variable should be assigning control and treatment sharply, there could be manipulation due to strategic layoffs. Citino et al. (2019) analyze the UI insurance program in place before ASpI, which had a 4 months UI benefit length discontinuity at age 50, and find evidence that 10% of layoffs in the two months prior to turning 50 were strategically delayed so as to obtain the longer UI benefit duration. While this finding pertains to the years 2010 to 2013 and to another program, it is not unlikely that a certain degree of manipulation in ASpI is going on and that our estimates are inflated.

7.2 Checking for Continuity of Baseline Covariates

We now turn to check whether the baseline covariates are continuous below and above the threshold. In the literature, when such tests are performed, it is with respect to continuous covariates around the threshold (see Canay and Kamat (2017) so-called induced ordered statistics permutation test). Unfortunately, our data set did not provide any continuous variable we could have used in such an analysis. Nevertheless, we observe five co-variates of which one is binomial and four are categorical. We have sex (SG11 = 1 for male, SG11 = 2 for female), marital status (STACIM = 1 if celibate, STACIM = 2 if married, STACIM = 3 if divorced, STACIM = 6 if widower), citizenship (CITTAD = 1 if Italian citizen, CITTAD = 2 if EU citizen, CITTAD = 3 if non-EU citizen), Italian macro-region of residence (RIP3 = 1 if living in Northern Italy, RIP3 = 2 if living in Center Italy, RIP3 = 3 if living in Southern Italy), and school degree (TISTUD = 1 if no degree, TISTUD = 2 if elementary degree only, TISTUD = 3 if middle school degree only, TISTUD = 4 if three years high school degree or professional degree, TISTUD = 5 if 5 years high-school degree, TISTUD = 6 if high-school music or arts degree, TISTUD = 7 if three years university degree in special needs universities, TISTUD = 8 if undergraduate degree, TISTUD = 9 if masters degree, TISTUD = 10 if undergraduate and masters degree single cycle). With these variables,

we can instead run a simple ad-hoc permutation test (due to the small sample size) of differences in proportions between groups of observations below and above the thresholds (ages 49 and 54) for both 2014 and 2015. Tables 17 to 20 in Section 8 show the results of such tests for 1000 permutations. We notice that the P values for all of the differences in proportions are significantly bigger than 0.05. Hence, we cannot reject the hypothesis that the proportions are the same for any co-variate at the 95% significance level[2]. Nevertheless, this means that we do not find any 'discontinuous' covariates at the age thresholds, and hence our SRD seems to be valid.

7.3 Bandwidth Sensibility Analysis

Finally, we check for the sensitivity of our estimate with respect to the selected bandwidth. To do so we turn to Tables 9 and 10.

Here, we have computed the LATE estimates using both half and double Imbens and Kalyanaraman (2012) risk-minimizing bandwidths. In the upper quadrant of Table 9 we can see that the optimal bandwidth total effect LATE estimate is 11.41, the half bandwidth 12.90, and the double bandwidth 10.42. For what concerns the lower quadrant of Table 9 the optimal bandwidth total effect LATE estimate is 30.63, the half bandwidth 22.20, and the double bandwidth 18.65. For what concerns the upper quadrant of Table 10 the optimal bandwidth total effect LATE estimate is 0.1455, the half bandwidth -1.1670, and the double bandwidth 4.4503. Finally, for what concerns the lower quadrant of Table 10 the optimal bandwidth total effect LATE estimate is 15.30, the half bandwidth 24.58, and the double bandwidth 13.08. As we can see our estimates are relatively sensitive to the bandwidth selection, with the discontinuity at age 55 in 2014 being the most sensitive estimate (the estimate for the half and double bandwidths is almost 10 months smaller than the one for the optimal bandwidth). Nevertheless, for what concerns the other discontinuities sensitivity issues do not seem to plague our results excessively. We conclude that while our estimates are fairly sensitive to bandwidth selection, they could be in the ballpark of the 'true' effect.

2 For what concerns the discontinuity at age 49 in 2014 we see that the P value associated with TISTUD = 5 is smaller than 0.05. This would indicate a difference in proportions of workers with at least 5 years of high school, and hence invalidate our identifying assumption that workers on the two sides of the cutoff only differ in their potential benefits duration. Since this is the only level whereby the P value is smaller than 0.05, and for the same level and the other discontinuities the P value is larger than 0.05 we interpret our positive finding to be driven by the poor nature of our data.

Table 9: SRD Bandwidth Sensitivity Analysis - ASpI 2014 Data (No Outliers)

ASpI 2014 Data			
	Discontinuity at Age 49		
LATE (DURRIC)	IK-Bw	Half-Bw	Double-Bw
Total Effect (Months)	11.41	12.90	10.42
	(8.170)	(11.837)	(6.761)
Week-per-Week	0.656	0.742	0.599
	(0.470)	(0.681)	(0.389)
Bandwidth	6.229	3.115	12.459
Observations	54	37	110
F Statistic	1.638	2.720	2.609
	(df = 3; 50)	(df = 3; 33)	(df = 3; 106)
	Discontinuity at Age 54		
LATE (DURRIC)	IK-Bw	Half-Bw	Double-Bw
Total Effect (Months)	30.63**	22.20**	18.65*
	(13.784)	(9.909)	(10.982)
Week-per-Week	3.524**	2.554**	2.146*
	(1.586)	(1.140)	(1.264)
Bandwidth	2.221	1.110	4.442
Observations	17	7	31
F Statistic	2.884	3.976	2.606
	(df = 3; 13)	(df = 1; 5)	(df = 3; 27)
			*p<0.1; **p<0.05; ***p<0.01

Note: Results of the SRD design using Imbens and Kalyanaraman (2012) formula for bandwidth, half-bandwidth and double-bandwidth selection. Regressing job-search durations (DURRIC) on age (ETAM) for the discontinuities in the ASpI 2014 program at ages 49 and 54. Using a triangular kernel and heteroskedasticity-consistent standard errors. Reporting both the total effect in months and the week-per-week effect of an increase in potential benefits duration. The week-per-week effect is calculated by dividing the full effect by the 'extension' in potential benefits duration corresponding to the age discontinuity and by the number of weeks in a month (4.34524). Removing job-search durations shorter than 48 months (4 years). Standard errors in parentheses.

Table 9: SRD Bandwidth Sensitivity Analysis - ASpI 2015 Data (No Outliers)

ASpI 2015 Data			
	Discontinuity at Age 49		
LATE (DURRIC)	IK-Bw	Half-Bw	Double-Bw
Total Effect (Months)	0.1455	-1.1670	4.4503
	(6.362)	(7.197)	(5.483)
Week-per-Week	0.018	-0.134	0.512
	(0.732)	(0.828)	(0.631)
Bandwidth	8.992	4.496	17.983
Observations	75	42	151
F Statistic	1.7864	0.5873	3.2624
	(df = 3; 71)	(df = 3; 38)	(df = 3; 147)
	Discontinuity at Age 54		
LATE (DURRIC)	Ik-Bw	Half-Bw	Double-Bw
Total Effect (Months)	15.30**	24.58***	13.08**
	(7.686)	(8.624)	(6.619)
Week-per-Week	0.88**	1.414***	0.752
	(0.442)	(0.496)	(0.380)
Bandwidth	4.561	2.280	9.122
Observations	23	11	29
F Statistic	0.7987	1.7390	0.6085
	(df = 3; 19)	(df = 1; 7)	(df = 3; 25)
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: Results of the SRD design using Imbens and Kalyanaraman (2012) formula for bandwidth, half-bandwidth and double-bandwidth selection. Regressing job-search durations (DURRIC) on age (ETAM) for the discontinuities in the ASpI 2015 program at ages 49 and 54. Using a triangular kernel and heteroskedasticity-consistent standard errors. Reporting both the total effect in months and the week-per-week effect of an increase in potential benefits duration. The week-per-week effect is calculated by dividing the full effect by the 'extension' in potential benefits duration corresponding to the age discontinuity and by the number of weeks in a month (4.34524). Removing job-search durations shorter than 48 months (4 years). Standard errors in parentheses.

8 Discussion and Conclusion

In this paper, we have analyzed the Italian UI benefits program *Assicurazione sociale per l'impiego* or ASpI for the years 2014 and 2015. More in particular we looked at the discontinuity in the length of UI benefit eligibility at ages 49 and 54. For the discontinuity at age 49, we obtained week-per-week results that contain previous estimates conducted in the Italian labor market Citino et al. (2019). For the discontinuity at age 54, we found signifi cantly larger estimates. We argued that such inflated values could be due to the fact that longer potential benefit durations at age 54 could offer a bridge to early retirement. Another source of upward bias that we did not account for in the paper is the possible sizeable ma nipulation of the forcing variable found by Citino et al. (2019) in the UI benefits program running between 2010 and 2013. Italian workers would collude with employers, strategically manipulating the layoff date to obtain a longer benefits duration.

We also analyzed the presence of a spike in exits from unemployment around benefits exhaustion, without being able to find clear evidence for Hypothesis 2. Nonetheless, our data severely limits our analysis. As has been noted in the paper the number of observations for the group g_{3j} for years $j \in (2014, 2015)$ was less than 10.

Our analysis is still interesting inasmuch as it considers a UI program upon which little analysis has been carried out and it provides tentative evidence concerning the direction of the causal effect, even if it cannot provide a reliable magnitude of the effect and if it cannot rule out a non-finding. Hence, it provides tentative evidence in support of Hypothesis 1 and gives resonance to the prediction of the flagship search-theoretic models. Our analysis could be better carried out with administrative data (see Caldura (2019) for a complete account of an easily accessible comprehensive Italian regional data source), which would eliminate reporting bias, and allow us to work with bigger sample sizes. Finally, further research should focus not only on the disincentive effects of the UI system but also on the consumption smoothing and job-match quality qualities of the ASpI program. Finally, a comprehensive analysis of the *Nuova assicurazione sociale per l'impiego* (NASpI) UI program, which since 2016 has been the main UI insurance instrument in Italy. While no age discontinuity in potential benefits duration is present for this program, the maximum benefits formula allows for identification using a Regression Kink Desing (RKD) *à la* Card et al. (2015).

References

Atkinson, A., Gomulka, J., Micklewright, J., and Rau, N. (1984). Unemployment bene fit, duration and incentives in britain: How robust is the evidence? *Journal of Public Economics*, 23(1):3–26.

Ben-Horim, M. and Zuckerman, D. (1987). The effect of unemployment insurance on unem ployment duration. *Journal of Labor Economics*, 5(3):386–390.

Bowman, D., Minehart, D., and Rabin, M. (1999). Loss aversion in a consumption–savings model. *Journal of Economic Behavior and Organization*, 38(2):155–178.

Caldura, F. R. M. (2019). Trasformazioni nel mercato del lavoro e nuovi approcci analitici i dati amministrativi del veneto: Utilizzo, confronto, analisi.

Canay, I. A. and Kamat, V. (2017). Approximate permutation tests and induced or der statistics in the regression discontinuity design. *The Review of Economic Studies*, 85(3):1577–1608.

Card, D., Chetty, R., and Weber, A. (2007a). Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market*. *The Quarterly Journal of Economics*, 122(4):1511–1560.

Card, D., Chetty, R., and Weber, A. (2007b). The spike at benefit exhaustion: Leaving the unemployment system or starting a new job? *American Economic Review*, 97(2):113–118.

Card, D., Lee, D. S., Pei, Z., and Weber, A. (2015). Inference on causal effects in a generalized regression kink design. *Econometrica*, 83(6):2453–2483.

Card, D. and Levine, P. B. (2000). Extended benefits and the duration of ui spells: evidence from the new jersey extended benefit program. *Journal of Public Economics*, 78(1):107– 138. Proceedings of the Trans Atlantic Public Economics Seminar on.

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal: Promoting communications on statistics and Stata*, 18(1):234–261.

Cattaneo, M. D., Jansson, M., and Ma, X. (2022). lpdensity: Local polynomial density estimation and inference. *Journal of Statistical Software*, 101(2).

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records *. *The Quarterly Journal of Economics*, 126(2):749–804.

Citino, L., Russ, K., and Scrutinio, V. (2019). Happy birthday? manipulation and selection in unemployment insurance.

DellaVigna, S., Lindner, A., Reizer, B., and Schmieder, J. F. (2017). Reference-Dependent Job Search: Evidence from Hungary*. *The Quarterly Journal of Economics*, 132(4):1969– 2018.

Efron, B. (1988). Logistic regression, survival analysis, and the kaplan-meier curve. *Journal of the American Statistical Association*, 83(402):414–425.

Fan, J. and Gijbels, I. (1994). Applications of local polynomial modelling. *Local Polynomial Modelling and its Applications*, page 159–216.

Giorgi, F. (2018). La recente evoluzione dell'indennità di disoccupazione in italia (the recent evolution of unemployment benefits in italy). *SSRN Electronic Journal*.

Greenwood, M. (1927). A report on the natural duration of cancer. JAMA: *The Journal of the American Medical Association*, 88(7):507.

Hunt, J. (1995). The effect of unemployment compensation on unemployment duration in germany. *Journal of Labor Economics*, 13(1):88–120.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959.

Imbens, G. and Lemieux, T. (2007). Regression discontinuity designs: A guide to practice. *NBER Working Paper Series*.

Inderbitzin, L., Staubli, S., and Zweimüller, J. (2016). Extended unemployment benefits and early retirement: Program complementarity and program substitution. *American Economic Journal: Economic Policy*, 8(1):253–88.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Katz, L. F. and Meyer, B. D. (1990a). The impact of the potential duration of unemployment benefits on the duration of unemployment. *Journal of Public Economics*, 41(1):45–72.

Katz, L. F. and Meyer, B. D. (1990b). Unemployment insurance, recall expectations, and unemployment outcomes. *The Quarterly Journal of Economics*, 105(4):973.

Kleven, H. J. and Waseem, M. (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan *. *The Quarterly Journal of Economics*, 128(2):669–723.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Lalive, R. (2008). How do extended benefits affect unemployment duration? a regression discontinuity approach. *Journal of Econometrics*, 142(2):785–806. The regression discontinuity design: Theory and applications.

Lalive, R. and Zweimüller, J. (2004). Benefit entitlement and unemployment duration: The role of policy endogeneity. *Journal of Public Economics*, 88(12):2587–2616.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Lentz, R. and Tranæs, T. (2005). Job search and savings: Wealth effects and duration dependence. *Journal of Labor Economics*, 23(3):467–489.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58(4):757–782.

Moffitt, R. (1985). Unemployment insurance and the distribution of unemployment spells. *Journal of Econometrics*, 28(1):85–101.

Moffitt, R. and Nicholson, W. (1982). The effect of unemployment insurance on unemployment: The case of federal supplemental benefits. *The Review of Economics and Statistics*, 64(1):1–11.

Mortensen, D. T. (1977). Unemployment insurance and job search decisions. *Industrial Labor Relations Review*, 30(4):505–517.

Mortensen, D. T. (1986). Chapter 15 job search and labor market analysis. *volume 2 of Handbook of Labor Economics*, pages 849–919. Elsevier.

Nekoei, A. and Weber, A. (2017). Does extending unemployment benefits improve job quality? *American Economic Review*, 107(2):527–61.

OECD (2015). Pensions at a glance 2015. *OECD Pensions at a Glance*.

Porter, J. (2003). Estimation in the regression discontinuity model. *Unpublished*.

Rogers, C. L. (1998). Expectations of unemployment insurance and unemployment duration. *Journal of Labor Economics*, 16(3):630–666.

Rogerson, R., Shimer, R., and Wright, R. (2005). Search-theoretic models of the labor market: A survey. *Journal of Economic Literature*, 43(4):959–988.

Rosolia, A. and Sestito, P. (2012). The effects of unemployment benefits in Italy: evidence from an institutional change. Temi di discussione (Economic working papers) 860, Bank of Italy, Economic Research and International Relations Area.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.

Schmieder, J. F., von Wachter, T., and Bender, S. (2016). The effect of unemployment benefits and nonemployment durations on wages. *American Economic Review*, 106(3):739– 77.

Scrutinio, V. (2019). The medium-term effects of unemployment benefits. *WorkINPS papers n. 18*.

Singer, J. D. and Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18(2):155– 195.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69(3):213– 225.

Tatsiramos, K. (2007). The effect of job displacement on the transitions to employment and early retirement for older workers in four european countries.

van den Berg, G. J. (1990). Nonstationarity in Job Search Theory. *The Review of Economic Studies*, 57(2):255–277.

van Ours, J. C. and Vodopivec, M. (2006). How shortening the potential duration of unem ployment benefits affects the duration of unemployment: Evidence from a natural experi ment. *Journal of Labor Economics*, 24(2):351–378.

Vodopivec, M. (2016). Unemployment insurance and duration of unemployment: Evidence from slovenia's transition.

Winter-Ebmer, R. (1998). Potential unemployment benefit duration and spell length: Lessons from a quasi-experiment in austria. *Oxford Bulletin of Economics and Statistics*, 60(1):33–45.

Appendix A

Table 11: Lifetable Group 1 - ASpI 2014

t_i	At Risk	Exits	Lost	$\hat{\lambda}(t_i)$	$\hat{\mathbb{V}}(\hat{\lambda}(t_i))$	95% CI $\hat{\lambda}(t_i)$	$\hat{S}(t_i)$	$\hat{\mathbb{V}}(\hat{S}(t_i))$	95% CI $\hat{S}(t_i)$
1	23	1	0	0.04	0.04	[-0.04,0.13]	0.96	0.04	[0.87,1.04]
2	22	5	0	0.23	0.18	[0.05,0.4]	0.74	0.19	[0.56,0.92]
3	22	5	0	0.23	0.18	[0.05,0.4]	0.74	0.19	[0.56,0.92]
4	17	2	0	0.12	0.14	[-0.04,0.27]	0.65	0.23	[0.46,0.85]
5	15	1	0	0.07	0.1	[-0.06,0.19]	0.61	0.24	[0.41,0.81]
6	14	1	0	0.07	0.11	[-0.06,0.21]	0.57	0.25	[0.36,0.77]
7	14	1	0	0.07	0.11	[-0.06,0.21]	0.57	0.25	[0.36,0.77]
8	13	1	0	0.08	0.13	[-0.07,0.22]	0.52	0.25	[0.32,0.73]
9	13	1	0	0.08	0.13	[-0.07,0.22]	0.52	0.25	[0.32,0.73]
10	13	1	0	0.08	0.13	[-0.07,0.22]	0.52	0.25	[0.32,0.73]
11	13	1	0	0.08	0.13	[-0.07,0.22]	0.52	0.25	[0.32,0.73]
12	12	5	0	0.42	0.47	[0.14,0.7]	0.3	0.21	[0.12,0.49]
13	12	5	0	0.42	0.47	[0.14,0.7]	0.3	0.21	[0.12,0.49]
14	12	5	0	0.42	0.47	[0.14,0.7]	0.3	0.21	[0.12,0.49]
15	12	5	0	0.42	0.47	[0.14,0.7]	0.3	0.21	[0.12,0.49]
16	12	5	0	0.42	0.47	[0.14,0.7]	0.3	0.21	[0.12,0.49]
17	12	5	0	0.42	0.47	[0.14,0.7]	0.3	0.21	[0.12,0.49]
18	12	5	0	0.42	0.47	[0.14,0.7]	0.3	0.21	[0.12,0.49]
19	12	5	0	0.42	0.47	[0.14,0.7]	0.3	0.21	[0.12,0.49]
20	7	0	7	0	0	[0,0]	0.3	0.21	[0.12,0.49]

Note: Life-table computing the empirical re-employment hazard and survival functions of Group 1 workers (age between 45 and 49) for the year 2014. 95% confidence intervals computed using Greenwood (1927) formula.

Table 12: Lifetable Group 2 - ASpI 2014

t_i	At Risk	Exits	Lost	$\hat{\lambda}(t_i)$	$\hat{\mathbb{V}}(\hat{\lambda}(t_i))$	95% CI $\hat{\lambda}(t_i)$	$\hat{S}(t_i)$	$\hat{\mathbb{V}}(\hat{S}(t_i))$	95% CI $\hat{S}(t_i)$
1	25	1	0	0.04	0.04	[-0.04,0.12]	0.96	0.04	[0.88,1.04]
2	25	1	0	0.04	0.04	[-0.04,0.12]	0.96	0.04	[0.88,1.04]
3	25	1	0	0.04	0.04	[-0.04,0.12]	0.96	0.04	[0.88,1.04]
4	24	1	0	0.04	0.04	[-0.04,0.12]	0.92	0.07	[0.81,1.03]
5	23	1	0	0.04	0.05	[-0.04,0.13]	0.88	0.11	[0.75,1.01]
6	22	2	0	0.09	0.09	[-0.03,0.21]	0.8	0.16	[0.64,0.96]
7	22	2	0	0.09	0.09	[-0.03,0.21]	0.8	0.16	[0.64,0.96]
8	22	2	0	0.09	0.09	[-0.03,0.21]	0.8	0.16	[0.64,0.96]
9	22	2	0	0.09	0.09	[-0.03,0.21]	0.8	0.16	[0.64,0.96]
10	22	2	0	0.09	0.09	[-0.03,0.21]	0.8	0.16	[0.64,0.96]
11	20	1	0	0.05	0.06	[-0.05,0.15]	0.76	0.18	[0.59,0.93]
12	19	4	0	0.21	0.22	[0.03,0.39]	0.6	0.24	[0.41,0.79]
13	19	4	0	0.21	0.22	[0.03,0.39]	0.6	0.24	[0.41,0.79]
14	19	4	0	0.21	0.22	[0.03,0.39]	0.6	0.24	[0.41,0.79]
15	19	4	0	0.21	0.22	[0.03,0.39]	0.6	0.24	[0.41,0.79]
16	19	4	0	0.21	0.22	[0.03,0.39]	0.6	0.24	[0.41,0.79]
17	19	4	0	0.21	0.22	[0.03,0.39]	0.6	0.24	[0.41,0.79]
18	15	1	0	0.07	0.1	[-0.06,0.19]	0.56	0.25	[0.37,0.75]
19	15	1	0	0.07	0.1	[-0.06,0.19]	0.56	0.25	[0.37,0.75]
20	14	1	13	0.07	0.12	[-0.06,0.21]	0.52	0.25	[0.32,0.72]

Note: Life-table computing the empirical re-employment hazard and survival functions of Group 2 workers (age between 50 and 54) for the year 2014. 95% confidence intervals computed using Greenwood (1927) formula.

Table 13: Lifetable Group 3 - ASpI 2014

t_i	At Risk	Exits	Lost	$\hat{\lambda}(t_i)$	$\hat{\mathbb{V}}(\hat{\lambda}(t_i))$	95% CI $\hat{\lambda}(t_i)$	$\hat{S}(t_i)$	$\hat{\mathbb{V}}(\hat{S}(t_i))$	95% CI $\hat{S}(t_i)$
1	13	1	0	0.08	0.07	[-0.07,0.22]	0.92	0.07	[0.78,1.07]
2	13	1	0	0.08	0.07	[-0.07,0.22]	0.92	0.07	[0.78,1.07]
3	12	1	0	0.08	0.08	[-0.07,0.24]	0.85	0.13	[0.65,1.04]
4	12	1	0	0.08	0.08	[-0.07,0.24]	0.85	0.13	[0.65,1.04]
5	12	1	0	0.08	0.08	[-0.07,0.24]	0.85	0.13	[0.65,1.04]
6	12	1	0	0.08	0.08	[-0.07,0.24]	0.85	0.13	[0.65,1.04]
7	12	1	0	0.08	0.08	[-0.07,0.24]	0.85	0.13	[0.65,1.04]
8	12	1	0	0.08	0.08	[-0.07,0.24]	0.85	0.13	[0.65,1.04]
9	12	1	0	0.08	0.08	[-0.07,0.24]	0.85	0.13	[0.65,1.04]
10	12	1	0	0.08	0.08	[-0.07,0.24]	0.85	0.13	[0.65,1.04]
11	11	1	0	0.09	0.1	[-0.08,0.26]	0.77	0.18	[0.54,1]
12	10	2	0	0.2	0.21	[-0.05,0.45]	0.62	0.24	[0.35,0.88]
13	10	2	0	0.2	0.21	[-0.05,0.45]	0.62	0.24	[0.35,0.88]
14	10	2	0	0.2	0.21	[-0.05,0.45]	0.62	0.24	[0.35,0.88]
15	10	2	0	0.2	0.21	[-0.05,0.45]	0.62	0.24	[0.35,0.88]
16	10	2	0	0.2	0.21	[-0.05,0.45]	0.62	0.24	[0.35,0.88]
17	10	2	0	0.2	0.21	[-0.05,0.45]	0.62	0.24	[0.35,0.88]
18	10	2	0	0.2	0.21	[-0.05,0.45]	0.62	0.24	[0.35,0.88]
19	10	2	0	0.2	0.21	[-0.05,0.45]	0.62	0.24	[0.35,0.88]
20	8	0	8	0	0	[0,0]	0.62	0.24	[0.35,0.88]

Note: Life-table computing the empirical re-employment hazard and survival functions of Group 1 workers (age between 55 and 59) for the year 2014. 95% confidence intervals computed using Greenwood (1927) formula.

Table 14: Lifetable Group 1 - ASpI 2015

t_i	At Risk	Exits	Lost	$\hat{\lambda}(t_i)$	$\hat{\mathbb{V}}(\hat{\lambda}(t_i))$	95% CI $\hat{\lambda}(t_i)$	$\hat{S}(t_i)$	$\hat{\mathbb{V}}(\hat{S}(t_i))$	95% CI $\hat{S}(t_i)$
1	34	1	0	0.03	0.03	[-0.03,0.09]	0.97	0.03	[0.91,1.03]
2	33	3	0	0.09	0.09	[-0.01,0.19]	0.88	0.1	[0.77,0.99]
3	30	1	0	0.03	0.04	[-0.03,0.1]	0.85	0.13	[0.73,0.97]
4	29	2	0	0.07	0.08	[-0.02,0.16]	0.79	0.16	[0.66,0.93]
5	29	2	0	0.07	0.08	[-0.02,0.16]	0.79	0.16	[0.66,0.93]
6	27	1	0	0.04	0.04	[-0.03,0.11]	0.76	0.18	[0.62,0.91]
7	27	1	0	0.04	0.04	[-0.03,0.11]	0.76	0.18	[0.62,0.91]
8	26	1	0	0.04	0.05	[-0.04,0.11]	0.74	0.19	[0.59,0.88]
9	26	1	0	0.04	0.05	[-0.04,0.11]	0.74	0.19	[0.59,0.88]
10	25	1	0	0.04	0.05	[-0.04,0.12]	0.71	0.21	[0.55,0.86]
11	25	1	0	0.04	0.05	[-0.04,0.12]	0.71	0.21	[0.55,0.86]
12	24	9	0	0.38	0.33	[0.18,0.57]	0.44	0.25	[0.27,0.61]
13	24	9	0	0.38	0.33	[0.18,0.57]	0.44	0.25	[0.27,0.61]
14	24	9	0	0.38	0.33	[0.18,0.57]	0.44	0.25	[0.27,0.61]
15	24	9	0	0.38	0.33	[0.18,0.57]	0.44	0.25	[0.27,0.61]
16	24	9	0	0.38	0.33	[0.18,0.57]	0.44	0.25	[0.27,0.61]
17	24	9	0	0.38	0.33	[0.18,0.57]	0.44	0.25	[0.27,0.61]
18	15	2	0	0.13	0.26	[-0.04,0.31]	0.38	0.24	[0.22,0.55]
19	15	2	0	0.13	0.26	[-0.04,0.31]	0.38	0.24	[0.22,0.55]
20	13	0	13	0	0	[0,0]	0.38	0.24	[0.22,0.55]

Note: Life-table computing the empirical re-employment hazard and survival functions of Group 1 workers (age between 45 and 49) for the year 2015. 95% confidence intervals computed using Greenwood (1927) formula.

Table 15: Lifetable Group 2 - ASpI 2015

t_i	At Risk	Exits	Lost	$\hat{\lambda}(t_i)$	$\hat{\mathbb{V}}(\hat{\lambda}(t_i))$	95% CI $\hat{\lambda}(t_i)$	$\hat{S}(t_i)$	$\hat{\mathbb{V}}(\hat{S}(t_i))$	95% CI $\hat{S}(t_i)$
1	18	1	0	0.06	0.05	[-0.05,0.16]	0.94	0.05	[0.84,1.05]
2	17	3	0	0.18	0.15	[0,0.36]	0.78	0.17	[0.59,0.97]
3	17	3	0	0.18	0.15	[0,0.36]	0.78	0.17	[0.59,0.97]
4	17	3	0	0.18	0.15	[0,0.36]	0.78	0.17	[0.59,0.97]
5	17	3	0	0.18	0.15	[0,0.36]	0.78	0.17	[0.59,0.97]
6	14	1	0	0.07	0.09	[-0.06,0.21]	0.72	0.2	[0.52,0.93]
7	14	1	0	0.07	0.09	[-0.06,0.21]	0.72	0.2	[0.52,0.93]
8	14	1	0	0.07	0.09	[-0.06,0.21]	0.72	0.2	[0.52,0.93]
9	13	1	0	0.08	0.1	[-0.07,0.22]	0.67	0.22	[0.45,0.88]
10	13	1	0	0.08	0.1	[-0.07,0.22]	0.67	0.22	[0.45,0.88]
11	13	1	0	0.08	0.1	[-0.07,0.22]	0.67	0.22	[0.45,0.88]
12	12	2	0	0.17	0.21	[-0.04,0.38]	0.56	0.25	[0.33,0.79]
13	12	2	0	0.17	0.21	[-0.04,0.38]	0.56	0.25	[0.33,0.79]
14	12	2	0	0.17	0.21	[-0.04,0.38]	0.56	0.25	[0.33,0.79]
15	12	2	0	0.17	0.21	[-0.04,0.38]	0.56	0.25	[0.33,0.79]
16	12	2	0	0.17	0.21	[-0.04,0.38]	0.56	0.25	[0.33,0.79]
17	12	2	0	0.17	0.21	[-0.04,0.38]	0.56	0.25	[0.33,0.79]
18	12	2	0	0.17	0.21	[-0.04,0.38]	0.56	0.25	[0.33,0.79]
19	12	2	0	0.17	0.21	[-0.04,0.38]	0.56	0.25	[0.33,0.79]
20	10	1	9	0.1	0.16	[-0.09,0.29]	0.5	0.25	[0.27,0.73]

Note: Life-table computing the empirical re-employment hazard and survival functions of Group 2 workers (age between 50 and 54) for the year 2015. 95% confidence intervals computed using Greenwood (1927) formula.

Table 16: Lifetable Group 3 - ASpI 2015

t_i	At Risk	Exits	Lost	$\hat{\lambda}(t_i)$	$\hat{\mathbb{V}}(\hat{\lambda}(t_i))$	95% CI $\hat{\lambda}(t_i)$	$\hat{S}(t_i)$	$\hat{\mathbb{V}}(\hat{S}(t_i))$	95% CI $\hat{S}(t_i)$
1	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
2	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
3	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
4	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
5	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
6	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
7	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
8	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
9	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
10	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
11	8	1	0	0.12	0.11	[-0.1,0.35]	0.88	0.11	[0.65,1.1]
12	7	3	0	0.43	0.28	[0.06,0.8]	0.5	0.25	[0.15,0.85]
13	7	3	0	0.43	0.28	[0.06,0.8]	0.5	0.25	[0.15,0.85]
14	7	3	0	0.43	0.28	[0.06,0.8]	0.5	0.25	[0.15,0.85]
15	7	3	0	0.43	0.28	[0.06,0.8]	0.5	0.25	[0.15,0.85]
16	7	3	0	0.43	0.28	[0.06,0.8]	0.5	0.25	[0.15,0.85]
17	7	3	0	0.43	0.28	[0.06,0.8]	0.5	0.25	[0.15,0.85]
18	7	3	0	0.43	0.28	[0.06,0.8]	0.5	0.25	[0.15,0.85]
19	7	3	0	0.43	0.28	[0.06,0.8]	0.5	0.25	[0.15,0.85]
20	4	0	4	0	0	[0,0]	0.5	0.25	[0.15,0.85]

Note: Life-table computing the empirical re-employment hazard and survival functions of Group 1 workers (age between 55 and 59) for the year 2015. 95% confidence intervals computed using Greenwood (1927) formula.Table 17: Check of Continuity of Baseline Covariates (Age = 49) - ASpI 2014

Co-variate	Observed Statistics	P value
Proportion of (SG11=1)	-0.042	0.755
Proportion of (SG11=2)	0.042	0.482
Proportion of (STACIM=1)	0.175	0.118
Proportion of (STACIM=2)	-0.036	0.69
Proportion of (STACIM=3)	-0.032	0.754
Proportion of (STACIM=6)	-0.107	1
Proportion of (CITTAD=1)	0.026	0.616
Proportion of (CITTAD=2)	-0.036	1
Proportion of (CITTAD=3)	0.01	0.672
Proportion of (RIP3=1)	-0.11	0.833
Proportion of (RIP3=2)	-0.039	0.728
Proportion of (RIP3=3)	0.149	0.199
Proportion of (TISTUD=1)	0	1
Proportion of (TISTUD=2)	-0.071	1
Proportion of (TISTUD=3)	-0.081	0.799
Proportion of (TISTUD=4)	-0.026	0.837
Proportion of (TISTUD=5)	0.312	0.023
Proportion of (TISTUD=6)	-0.071	1
Proportion of (TISTUD=7)	-0.026	0.851
Proportion of (TISTUD=8)	0	1
Proportion of (TISTUD=9)	0	1
Proportion of (TISTUD=10)	-0.036	1

Note: Checking the 'continuity' of baseline co-variates using an ad-hoc permutation test of proportion difference for the discontinuity at age 49 in 2014. Performing 1000 permutations.

Table 18: Check of Continuity of Baseline Covariates (Age = 54) - ASpI 2014

Co-variate	Observed Statistics	P value
Proportion of (SG11=1)	0.012	0.612
Proportion of (SG11=2)	-0.012	0.7
Proportion of (STACIM=1)	0.032	0.663
Proportion of (STACIM=2)	-0.131	0.862
Proportion of (STACIM=3)	-0.008	0.714
Proportion of (STACIM=6)	0.107	0.44
Proportion of (CITTAD=1)	-0.071	1
Proportion of (CITTAD=2)	0.036	0.748
Proportion of (CITTAD=3)	0.036	0.754
Proportion of (RIP3=1)	-0.238	0.95
Proportion of (RIP3=2)	0.357	0.045
Proportion of (RIP3=3)	-0.119	0.865
Proportion of (TISTUD=1)	0	1
Proportion of (TISTUD=2)	-0.04	0.852
Proportion of (TISTUD=3)	-0.242	0.951
Proportion of (TISTUD=4)	-0.04	0.864
Proportion of (TISTUD=5)	0.143	0.287
Proportion of (TISTUD=6)	0.071	0.575
Proportion of (TISTUD=7)	0.071	0.577
Proportion of (TISTUD=8)	0	1
Proportion of (TISTUD=9)	0	1
Proportion of (TISTUD=10)	0.036	0.787

Note: Checking the 'continuity' of baseline co-variates using an ad-hoc permutation test of proportion difference for the discontinuity at age 54 in 2014. Performing 1000 permutations.

Table 19: Check of Continuity of Baseline Covariates (Age = 49) - ASpI 2015

Co-variate	Observed Statistics	P value
Proportion of (SG11=1)	-0.044	0.728
Proportion of (SG11=2)	0.044	0.482
Proportion of (STACIM=1)	0.142	0.185
Proportion of (STACIM=2)	-0.094	0.846
Proportion of (STACIM=3)	-0.048	0.812
Proportion of (STACIM=6)	0	1
Proportion of (CITTAD=1)	0.068	0.402
Proportion of (CITTAD=2)	-0.009	0.744
Proportion of (CITTAD=3)	-0.059	0.885
Proportion of (RIP3=1)	-0.076	0.81
Proportion of (RIP3=2)	0.062	0.396
Proportion of (RIP3=3)	0.014	0.583
Proportion of (TISTUD=1)	0	1
Proportion of (TISTUD=2)	0.061	0.398
Proportion of (TISTUD=3)	-0.126	0.888
Proportion of (TISTUD=4)	-0.07	0.946
Proportion of (TISTUD=5)	0.244	0.057
Proportion of (TISTUD=6)	-0.05	1
Proportion of (TISTUD=7)	-0.05	1
Proportion of (TISTUD=8)	-0.05	1
Proportion of (TISTUD=9)	0	1
Proportion of (TISTUD=10)	0.041	0.529

Note: Checking the 'continuity' of baseline covariates using an ad-hoc permutation test of proportion difference for the discontinuity at age 49 in 2015. Performing 1000 permutations.

Table 20: Check of Continuity of Baseline Covariates (Age = 54) - ASpI 2014

Co-variate	Observed Statistics	P value
Proportion of (SG11=1)	0.094	0.442
Proportion of (SG11=2)	-0.094	0.795
Proportion of (STACIM=1)	0.1	0.469
Proportion of (STACIM=2)	-0.078	0.808
Proportion of (STACIM=3)	-0.022	0.735
Proportion of (STACIM=6)	0	1
Proportion of (CITTAD=1)	-0.139	0.918
Proportion of (CITTAD=2)	0.1	0.443
Proportion of (CITTAD=3)	0.039	0.628
Proportion of (RIP3=1)	-0.278	0.971
Proportion of (RIP3=2)	0.15	0.312
Proportion of (RIP3=3)	0.128	0.418
Proportion of (TISTUD=1)	0	1
Proportion of (TISTUD=2)	-0.222	1
Proportion of (TISTUD=3)	0.328	0.093
Proportion of (TISTUD=4)	-0.122	0.925
Proportion of (TISTUD=5)	-0.183	0.942
Proportion of (TISTUD=6)	0.05	0.673
Proportion of (TISTUD=7)	0.05	0.684
Proportion of (TISTUD=8)	0.05	0.686
Proportion of (TISTUD=9)	0	1
Proportion of (TISTUD=10)	0.05	0.673

Note: Checking the 'continuity' of baseline covariates using an ad-hoc permutation test of proportion difference for the discontinuity at age 54 in 2015. Performing 1000 permutations.

Appendix B

This section relies heavily on Giorgi (2018).

Before 2013

Before 2013 the Italian UI system consisted of three programs:

- ***Indennità di disoccupazione***

This instrument could be requested by involuntarily unemployed dependent workers (excluding apprentice workers) with at least 2 years of total contributions to social security and at least 52 weeks of contributions to social security in the 2 years prior to lay-off. The UI benefit duration consisted of 8 months for workers below 50 years of age and 12 months for workers above and including 50 years of age. The benefit amounted to 60% of the mean salary (calculated in the 3 months before lay-off) for the first 6 months of unemployment, 50% from the 6th month to the 8th month, and 40% until benefit expiration.

- ***Indennità a requisiti ridotti***
  This instrument could be requested by involuntarily unemployed dependent workers with at least 2 years of total contributions to social security and at least 11 weeks of contributions to social security in the 2 years prior to lay-off. The UI benefit duration consisted of 6 months. The benefit amounted to 35% of the mean salary (calculated in the 6 months before lay-off) for the first 4 months of unemployment, and 40% until benefit expiration.

- ***Mobilità***
  This instrument could be requested by involuntarily unemployed dependent workers with a so-called *contratto a tempo indeterminato* or a permanent contract with at least 12 months of company seniority, who had been subject to collective lay-offs by companies with 15 or more employees, or who were still unemployed once their redundancy fund (*Cassa Integrazione Straordinaria* or *CIGS*) access rights have expired. The duration of this instrument varied with geographical location (Northern, Central, and Southern Italy), and it varied year by year. The benefit amounted to 100% of the respective redundancy benefit amount for the first 12 months, and to 80% for the remaining period.

Between 2013 and 2015

In 2012, partly as a result of the ensuing European sovereign debt crisis, the Italian UI system was strongly reformed. The *LEGGE 28 giugno 2012, n. 92*, in effect on January 1st 2013, supplanted the *Indennità di disoccupazione* and the *Indennità a requisiti ridotti*, and introduced the *Assicurazione sociale per l'impiego* (ASpI) and the Mini-ASpI (the *Mobilità* instrument remained in place until 2016 included).

- ***Assicurazione sociale per l'impiego (ASpI)***
  This instrument could be requested by involuntarily unemployed dependent workers (including apprentice workers) with at least 2 years of total contributions to social security and at least 12 months of contributions to social security in the 2 years prior to lay-off. The UI benefit duration in 2013 consisted of 8 months for workers below 50 years of age, and 12 months for workers above and including 50 years of age. In 2014 the UI benefit duration consisted of 8 months for workers below 50 years of age, of 12 months for workers aged between 50 and 54, and 14 months for workers aged above and including 55 years of age. In 2015 the UI benefit duration consisted of 10 months for workers below 50 years of age, 12 months for workers aged between 50 and 54, and 16 months for workers aged above and including 55 years of age. The benefit amounted to 75% of the mean salary (calculated in the previous 2 years before lay-off) for the first 6 months of unemployment, with a 15% reduction at the 7th month and 13th month.

- ***Mini-ASpI***
  The instrument could be requested by involuntarily unemployed workers with short work experience. Similarly to the *Indennità a requisiti ridotti* instrument, it only required a 13 weeks contribution to social security in the 12 months before lay-off. It lasted for 50% of the weeks the unemployed worker contributed to social security in the 12 months previous to lay-off.

From 2016 onward

In 2015 a further normative change occurred (legislative decree *D.Lgs. 4 marzo 2015, n. 22.*), which unified the ASpI and and Mini-ASpI under the same program called *Nuova assicurazione sociale per l'impiego* (NASpI).

- ***Nuova assicurazione sociale per l'impiego (NASpI)***
  This instrument could be requested by involuntarily unemployed dependent workers (including apprentice workers) without total contributions to social security require ment, but with at least 13 weeks of contributions to social security in the 4 years prior to lay-off, and 30 work days in the 12 months prior to lay-off. The UI benefit duration stopped being dependent on age, but it consisted of 50% of the weeks of contribu tions to social security in the previous 4 years, with a max of 24 months. The benefit amounted to 75% of the mean salary (calculated in the previous 4 years before lay-off) for the first 3 months of unemployment, with a 3% reduction every month starting from the 4th month.

Measuring Monetary Policy Uncertainty in India

August 16, 2023

Abstract

Modelling domestic economic uncertainty is not only of crucial importance to individual investors but also relevant for policymakers to effectively set monetary policy measures, particu larly in emerging markets. This proposal presents a subcategorical monetary policy uncertainty index for India, following the widely used newspaper-based approach for economic policy un certainty. While strongly correlated, a comparison between Indian economic policy uncertainty and monetary policy uncertainty reveals interesting insights into the determining factors of of interest for monetary policy uncertainty. Particularly, the dominating impact of US monetary policy on Indian uncertainty is highlighted. Further, this paper provides a comprehensive em pirical framework to investigate how monthly macroeconomic variables react to a one-standard deviation shock of elevated monetary policy uncertainty in India. Applying a Vector Error Correction Model and estimating the impulse responses through bootstrapping, this study finds that macroeconomic series, in particular GDP and the broad real exchange rate, are significantly negatively affected by a one-standard-deviation increase of monetary policy uncertainty. This confirms that shocks in domestic monetary policy uncertainty negatively affect macroeconomic indicators.

Content

1 Introduction

The effect of economic uncertainty across the globe has received large attention in recent years. Accompanying the integration of the financial world and development of the recent financial crisis in 2008 or the Covid-19 crisis, the dynamic effects of uncertainty shocks across the economic and financial system have sparked great interest. Understanding in particular the facets of economic policy uncertainty (EPU) and transmissions of impact is more relevant than ever. Not only economic recessions, but also political or societal changes have resulted in record-high levels of uncertainty, which could be of concern for aggregate economic activity. Elevated economic uncertainty may have different causes, yet the uncertainty about monetary policy is a crucial factor during economic recessions. Reducing uncertainty about central bank's monetary policy actions has been an objective of the Federal Reserve (Fed) or the European Central Bank, among other central banks, through forward-guidance in the last years. Nevertheless, investigating the transmission of monetary policy uncertainty (MPU) shocks on macroeconomic series has retained a particular interest. In the realm of monetary policy, uncertainty is a "defining characteristic of [its] landscape", as Greenspan (2003) stressed. Hence, not only has elevated uncertainty undermining effects on the effectiveness of monetary policy tools, it can also have severe effects on economic growth and distort the dynamics of the economy.

Over the past two decades, international economic activity has rapidly changed. Financial markets have become increasingly internationally linked, as emerging markets are becoming an essential element of modern financial markets. Emerging markets, like Brazil, India, China, and South Africa (BRICS) have become of importance for a long time, due to their driving influence on global economic development and high economic growth rates. In particular, India's economy will continue to play a vital role in the future, as the World Bank projects a gross domestic product (GDP) growth of 8.3% for 2022-2023 in their annual Global Economic Prospects report (2022). Understanding India's monetary policy and uncertainty is crucial to stay abreast of its time-varying global integration into the global economy. Originating in the United States (US), the financial crisis of 2008 spread with remarkable speed across both developed and emerging economies, affecting financial and real markets alike (Chudik & Fratzscher, 2011). Thirteen years later, the Covid 19 virus led to an unprecedented global pandemic, halting international financial markets and driving economies into recessions. Despite India being one of the most severely affected countries by the outbreak of the pandemic, its

economy recovered at rapid speed to pre-pandemic output levels (Worldbank, (2022)). Researchers and policymakers recognise India's potential, however little is known about its domestic economic uncertainty, particularly the impact and consequences of India's MPU. Therefore, it appears of relevance to analyse India's monetary policy and the economic impacts of elevated uncertainty.

As shown by Bloom (2007), uncertainty not only increases after economic or political shocks but also has significant effects on output and employment. Baker, Bloom, and Davis (2016) discover similar results for the effect of EPU, measured by the frequency of certain terms in newspaper coverage. Continuing the research on the categorical MPU, Husted et al. (2020) also confirm the powerful effects of MPU in the US with similar results. However, little research has been done to model EPU and specifically MPU in the emerging market of India. For example, Bhagat et al. (2016) find significant evidence of negative relations between Indian EPU and GDP growth or stock market performance in India. Nonetheless, most literature is mainly focused on EPU when analysing shocks to uncertainty. Thus, this paper presents an investigation into the macroeconomic effects of Indian MPU and finds significant effects both domestically as well as for international spillovers.

This paper extends the results of Bhagat et al. (2016) and Ghosh et al. (2021) by constructing a new MPU index and examining whether uncertainty shocks are transmitted into macroeconomic perfor mance indicators. For this purpose, monthly data from April 2004 to August 2021 is obtained for four macroeconomic variables: GDP, wholesale price index, broad real exchange rate, and the policy repo rate. The impulse response functions (IRFs) of the vector error correction model (VECM) are estimated through bootstrapping to assure a more precise investigation of dynamic effects. The applied bootstrapping method was first developed by Benkwitz and L¨utkepohl (2001).

This paper contributes in several ways to existing literature. First, this paper is the first to construct a novel MPU index for India using newspaper coverage frequency. Second, this study provides an extensive analysis to further investigate the dynamic effects of uncertainty shocks driven by mon etary policy on the Indian economy. Lastly, possible permanent effects are even indicated, as the cointegration analysis of the considered macroeconomic series revealed cointegration on order one.
The results of this paper stress the importance of further investigation into the effects of MPU on the Indian economy. Significant negative ef

fects on GDP and the real exchange rate verify that an increase in MPU will cause negative domestic economic effects. The effect on GDP reaches a peak decline of 0.3%, which is similar to the response of Chinese output when confronted by an MPU shock, as evaluated by Huang and Luk (2020). However, inflation remains largely unaffected by a shock in MPU. This points to the slow-moving nature of inflation and the persistent inflation ary environment of the Indian economy. In addition, this observation was confirmed by variance decomposition as inflation is largely determined by itself. In the long run, output is also largely determined by itself, yet MPU does have some influence, especially in the short run. Another key result of this paper is the influence of the US financial markets on the Indian economy, as the fall in GDP growth is more severe when integrating the US stock market into the model.

The conclusions drawn from this paper underline not only the need for a profound understanding of India's monetary policy framework but also its effects on India's economy, which has not been researched extensively yet.

The paper is structured as follows: Section 2.1 presents a theoretical review of the current litera ture on microeconomic, as well as macroeconomic effects of uncertainty and how macroeconomic shocks are transmitted across markets. Furthermore, section 2.2 concentrates on understanding and reviewing the literature on India's domestic monetary policy. In section 3.1, a newly constructed index measuring MPU in India is introduced and compared to the existing EPU for Indian uncer tainty. Section 4 introduces the obtained data and discusses the employed data manipulation. The following sections 4.1 and 4.2 examine the data on stationarity and cointegration. Based on these results, section 5 illustrates the methodology and model specification used for the analysis, and finally the modelled impulse response analysis using bootstrapping. Finally, section 7 concludes with a critical discussion of the implications and limitations of the findings

2 Literature Review

This section reviews the literature on MPU and its economic effects considering the Indian monetary policy framework. First, Section 2.1 summarises theoretical discovery on modelling uncertainty and its macroeconomic effects. Second, an overview of monetary policy in India and its objectives are discussed in Section 2.2.

2.1 Modelling uncertainty and measuring its economic effects

Given its normative aspect, a clear definition of uncertainty has been difficult to construct from an economic perspective. However, Jurado et al. (2015, p. 1177) capture the essence of uncertainty in their definition as the "conditional volatility of a disturbance that is unforecastable from the perspective of economic agents". Unfortunately, existing proxies or indicators to measure uncertainty, like the volatility of stock market returns or survey-based forecasts, do not entirely capture the aspect of uncertainty and have left large room for development. This led to several attempts in the past to measure economic uncertainty, among which the most prominent index was developed by Baker et al. (2016). They propose a newspaper-based index to capture EPU for various countries. In this methodology, they apply newspaper article coverage frequency related to EPU by controlling for the initiator, actions and effects of economic policy actions through specific categorical terms. Recent empirical studies have extended the investigations of EPU by focusing on categorical in dices. For instance, Husted et al. (2020) developed an index measuring MPU for the US, while Chen and Tillmann (2021) focused their efforts on a Chinese MPU and its spillover impacts on neighbouring Asian economies. Arbatli et al. (2017) implemented the idea of categorical indices in their analysis of policy uncertainty in Japan and Hardouvelis et al. (2018) constructed the index including category-specific EPU indices for the Greek economy. A similar approach was taken for Turkey by Cevik and Erduman (2020). The heterogeneous nature of uncertainty shocks underlines the importance of the different types of EPU shocks (Trung, 2019; Gabauer & Gupta, 2018). By focusing on different structures of policy uncertainty, such as trade policy or monetary policy, the proximate sources of policy uncertainty allow a closer diagnosis of effects on dynamic effects of the macro-economy or firm-level outcomes.

Uncertainty has been identified to lead to economic consequences through multiple channels by the oretical academic literature. Bernanke (1983) finds the reduction or delay of firm-level investments to be significant after uncertainty shocks. Pastor and Veronesi (2013) confirm this, in addition to corporate investments, and point to the increase in risk premiums through financial frictions models. This, in turn, increases the cost of capital and reduces the overall investment. Besides the reduction of investment, Bloom (2014) further highlights that the recruitment of new employees falls sharply by heightened uncertainty. As a result of this, Leduc and Liu (2016) show that uncertainty shocks, in general, reduce demand on the consumer side. They go as far as proposing similar effects as an aggregate demand shock. Using DSGE models, Leduc and Lui (2016) point to labour market search frictions, amplifying the effects of uncertainty. In this sense, with firms reducing employ ment opportunities, this will lead to postponed consumption, which in turn also affects the price level. Furthermore, households are incentivised to postpone consumption (Eberly, 1994) and pre pone savings out of precautionary reasons and smaller prediction confidence (Fern´andez-Villaverde, Guerr´on-Quintana, Kuester, & Rubio-Ram´ırez, 2015). Considering the broader macroeconomic performance, Baker et al. (2016), Leduc and Lui (2016), Colombo (2013), and Stock and Watson (2012) emphasise the decline in macroeconomic performance by elevated uncertainty. Also, inter national trade in services or goods experiences negative effects (Constantinescu, Mattoo, & Ruta, 2017).

Another strand of literature tests these theoretical assumptions and investigates the effects of policy uncertainty on economic outcomes empirically. For the US, Baker et al. (2016) conclude that elevated levels of policy uncertainty lead to a decline in employment growth and investment rates, while at the same time stock price volatility is heightened. Negative effects on stock price volatility are confirmed among others by Bloom (2009), Zhang et al. (2019), Leduc and Lui (2016), Julio and Yook (2012), and Kelly et al. (2016), particularly in an uncertain environment following elections. Julio and Yook (2016) also identify adverse effects on cross-country capital flows and corporate investments (2012). Through a substantial effect on the widening of credit spreads, as well as a decrease in equity returns, financial markets are adversely impacted by uncertainty shocks (Popp & Zhang, 2016). Results from studies concerned with merger and acquisition activities, as well as initial public offerings indicate substantial negative effects during times of higher EPU ((Bonaime, Gulen, & Ion, 2018) and (C¸ olak, Durnev, & Qian, 2017) respectively). Most parts of academic literature have been focusing on the effects of EPU in developed countries, in particular, the US and European countries. Yet, only a few empirical works have turned to economic outcomes of uncertainty in emerging markets like India. Anand and Tulin (2014) support the existing evidence that Indian firms react with an increase in cancelled or postponed investments to heightened uncertainty. Analysing the effects of shocks on the Indian macroeconomy, Gosh et al. (2021) point to the distinct influence of global uncertainty on monetary policy decisions in India. With the use of a VAR-X model, they find a negative effect on expected inflation, inflation rate, and the output gap. Overall,

Bhagat et al. (2016) conclude that EPU plays a significant role in the negative effects on the Indian GDP and Indian fixed investments, given the magnitude of their results. Further, Xiao-lin Li et al. (2016) investigate the relationship between stock return and EPU in China and India but find a negative impact on stock returns primarily in sub-periods.

2.2 Monetary policy in India

Not only is the understanding of the Indian monetary policy framework important for the analysis of perceived uncertainty, but its fundamental modifications and developments to further global in tegration also deserve attention by its own merit. The Reserve Bank of India Act of (RBI, 1934, p. 14) provides the mandate of the Reserve Bank of India (RBI) to "to regulate the issue of Bank notes and keeping of reserves with a view to securing monetary stability in India and generally to operate the currency and credit system of the country to its advantage; to have a modern monetary policy framework to meet the challenge of an increasingly complex economy, to maintain price sta bility while keeping in mind the objective of growth." After achieving independence in 1947, India's primary objectives of monetary policy, as stated now in the preamble, changed and evolved until their final implementation in 2016.

Due to the increasing integration of India's economy into global markets, financial stability has also become a key objective (Hammond, Kanbur, & Prasad, 2009). Dua (2020) emphasises two critical developments in the RBI's monetary policy framework since 1998. On the one hand, the implementation of a multiple indicator approach (MIA) removed the dollar as the only anchor for monetary policy and included variables like credit, output, trade, repo rate, inflation rate, and the exchange rate, among others, to aid in the formulation of monetary policy. On the other hand, inflation targeting was officially adopted in 2012 with a 4% target level of inflation. This framework is adopted to increase independence from fiscal authorities and policies (Ghate & Kletzer, 2016). Ghate and Kletzer (2016) mention this development in their discussion of capital inflows in 2008- 2010. Moving from tight management of exchange rate and current account deficits to adjusted capital controls with a focus on tightened inward capital control, enabled the RBI to manoeuvre itself out of constrained monetary policy. While Ghosh et al. (2016) stipulate that the Global Financial Crisis was overcome with capital account liberalisation, monetary policy autonomy, and exchange rate sensibility, this integration in the international economy has also become a challenge.

Thus, effective monetary transmission through changes in policy instruments is essential to steer investment, consumption, and overall demand to achieve monetary policy goals of price stability and growth. Although the RBI has used conventional monetary policy measures[1] to reach its objective, it had to resort to multiple unconventional tools in the past. This could be due to the influence of exchange rate movements in the transmission of monetary policy, as stressed by Al-Mashat (2003). Singh and Kalirajan (2007), however, also underline the important role play of interest rates in monetary policy transmission. In particular, the mirroring movements of the Federal funds rate could indicate strong dependencies and constraint's on Indian monetary policy (Aleem, 2010). This is further confirmed by Eichgreen (2013), Banerjee et al. (2021), and Fratzscher et al. (2012), who warn about the US spillover policy effects on other markets, in particular emerging countries.

3 Measuring monetary policy uncertainty in India

The following section is concerned with the introduction of a new index of MPU for India based on newspaper coverage frequency. It aims to mirror the approach put forward by Baker et al. (2016), by observing the frequency of terms associated with monetary policy uncertainty in newspaper articles. Hence, uncertainty is correlated with a particular set of keywords, which are selected to capture a meaningful indication of MPU. It will also be compared to the existing EPU index for India by Baker et al. (2016) to highlight the differences between EPU and MPU.

3.1 Construction of the index

Following the approach by Baker et al. (2016) for their Indian EPU index, the MPU index will be based on English-speaking newspapers only and reflect uncertainty related to the policy by the RBI. In order to compare the MPU and EPU indices, the index does not include Hindi-speaking newspapers. Since foreign investors would face a language barrier and rely primarily on international newspapers, this approach avoids distortion in the true degree of uncertainty. Using the Access World News database, the primary newspaper sources for the construction of the index are The Economic Times, Times of India, Indian Express, The Hindu,

1 Refer to Mohan (2008), Mohan, (2006b), Hammond (2009), Mohan, (2006a) for the conventional transmission channels of monetary policy including the interest rate, the exchange rate, expectations, and credit aggregates respec tively.

The Financial Express, and Hindustan Times from January 2002 until November 2021. Although other newspapers could certainly also classify as necessary to consider, with the above mentioned six newspapers the length of computation of the index reached about three weeks. Hence, it was necessary to strike a balance between completeness and tractability of the index. To classify as relevant for MPU, at least one term of three categories must appear in the article. The categories for the terms are displayed in Table 1. The approach taken in this paper differs from other MPU indices as it retains Baker et al.'s (2016) first category of "economy". This is done to ensure the overall integration of articles in the broader economic context and to do justice to the categorisation of this index of broader EPU. However, the second category only contains keywords related to monetary policy and adopts similar terms as proposed by Arbatli et al. (2017) for their construction of the Japanese MPU index, since they include a broad range of terms concerning monetary policy. The last category follows the literature on newspaper-based indices and contains words related to uncertainty. A future line of thought could consider including the terms "Reserve Bank of India" or "RBI" as a separate necessary category. This could enhance the degree of purely domestic MPU by limiting spillovers that are caused by global uncertainty. By using this approach of selecting certain keywords, it is assumed that the overall uncertainty in the economy related to monetary policy issues is captured by newspaper reports. As mentioned above, the rapid dissemination and frequent reporting of any information through newspapers thus offer a valuable measure of MPU.

Table 1: Index keywords

Term - Economic	economy, financial
Term - Monetary Policy	monetary policy(ies), Reserve Bank of India, RBI, credit control, quantitative easing, monetary easing, monetary tightening, interest rate, policy rate, monetary operations, market operations, inflation target, price target, expansionary monetary policy, further easing,, official discount rate, contractionary monetary policy
Term - Uncertainty	uncertainty, uncertain, volatile, unstable, unclear, unpredictable

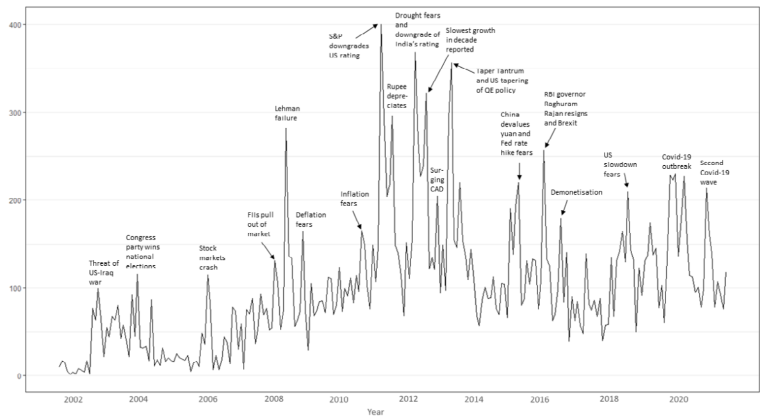
To control for differing article volume and coverage over time and newspaper, Baker et al. (2016) propose the following approach. Firstly, for each newspaper, the raw MPU counts are scaled by the total number of articles appearing per month to obtain a relative MPU frequency count. With this first step, any variations in the volume of articles for each of the six newspapers are accounted for. Secondly, the relative MPU frequency count is standardised per newspaper to the unit standard deviation from 2002 to 2021. As some newspapers have a higher coverage of monetary policy-related issues than others, this step limits the variability and ensures an equal weight of the newspapers in the index. As a third and last step, the resulting series is subsequently normalised to the mean of 100 across all six newspapers by month, resulting in the overall monthly Indian MPU index. Formally, the computation can be described as follows:

$i =$ newspaper; $t =$ time; $n =$ number of newspapers; $\sigma =$ unit standard deviation

$$X_{it} = \frac{\text{mpu count }_{it}}{\text{total }_{it}};$$
$$Y_{it} = \frac{X_{it}}{\sigma_i};$$
$$Z_t = \frac{\sum_i^n Y_{it}}{n};$$
$$M = \text{ Average } Z_t \text{ over } T;$$
$$\text{MPU}_t = \frac{Z_t}{M}$$

Figure 1 depicts the final Indian MPU index spanning a sample period of January 2002 to November 2021.

Figure 1: Indian Monetary Policy Uncertainty



This figure plots the Indian MPU from January 2002 until November 2021 obtained through the newspaper coverage approach. CAD denotes the abbreviation for Current Account Deficit, FIIs refers to Foreign Institutional Investors, and QE is short for Quantitative Easing.

A central limitation encountered when using the methodology described above is that one cannot capture the topics that were important during the relevant times. Higher MPU measured in one month might be caused by other concerns than in another month. Yet, to fully understand the economic and political developments that caused MPU, it is crucial to understand the topics that plagued the markets. More importantly, without such an understanding, certain trends or evolutions in macroeconomic indicators cannot be observed. For instance, inflation and the exchange rate have always been underlying concerns for the Indian market.

The overarching trend of increasing MPU in recent years goes hand-in-hand with India's growing exposure to global markets with continuing liberalisation efforts. The index peaked first in 2004 during the re-election of the Congress Party in National Elections in May. In the period between 2004 and 2008, inflationary concerns, a volatile rupee, increasing oil prices, and fear of rate hikes were among the most frequently mentioned topics. After initially assuming no impact of the US subprime mortgage crisis, the RBI reversed its judgement in September 2007 and predicted an economic slowdown. Afterwards, inflation continued to rise and rate hike efforts remained unsuccessful. The bankruptcy of the Lehman Brothers initiated the predicted recession in September 2008, causing the rupee to fall dramatically, the RBI to decrease the interest rates and introduce further measures in order to increase credit flow. Shortly after the Financial Crisis, the BSE Sensex returned to an upward trajectory, however, a negative inflation rate caused it to tumble again. 2009 and 2010 were primarily coined by rate hike fears and worrying about rising inflation rates. The later worry extended into the following years and continued to be a deepening concern for the RBI, which subsequently adjusted its policies giving priority to inflation control over growth aspirations. Thus, MPU levels did not return to pre-crisis levels and underlying uncertainty remained a key driver in the markets. While India returned to a positive growth trajectory after the financial crisis, systemic weaknesses have remained entrenched in the Indian economy (Reddy, 2010). MPU reached its highest peak after Standard & Poore's (S&P) downgraded the US rating to AA+ as a result of the Budget Control Act passing to end the US Debt Ceiling Crisis. Consequently, the rupee plunged and Moody's also downgraded Indian banks a months later. Over the entire period, it becomes clear that uncertainty about US monetary policy dominated India's MPU from 2013 until the beginning of 2015 and continued to be a large influence. In order to stop the free fall of the rupee, the RBI adopted a floating exchange rate regime in the beginning of 2012. Against the backdrop of economic development, the markets turned volatile in 2013 because of large capital outflows by Foreign Institutional Investors (FIIs), often referred to as Taper Tantrum. Furthermore, markets feared the imminent tapering of the Fed and consequential impacts on emerging markets. While the Greek crisis overshadowed the markets afterwards, MPU calmed until August 2015, when an unexpected devaluation of the yuan sent shock waves through the Indian market. Importantly, during this time the rupee among other emerging market currencies continued to follow its depreciating course.

Following unease about a potential economic slowdown in China, further domestic events caused the MPU to increase again. Of notable importance was June 2016, when the Brexit referendum on the 23rd of June coincided with the RBI governor Raghuram Rajan's announcement of retirement a few days before on the 18th of June (Rebello, 2016). Moreover, the decision by the government to demonetise all |500 and |1'000 banknotes in November 2016 affected MPU heavily and continued to be a topic over the next years. Finally, the last hikes in 2020 and 2021 were caused by uncertainty regarding the Covid-19 pandemic, lockdowns, but also cryptocurrencies. Particularly, this moderate effect of the Covid-19 crisis may seem puzzling at first. However, this trend is also in line with the results obtained by Iyke (2020) of a rather moderate impact on Indian EPU during the pandemic. An exhaustive listing of all key political and economic developments in uncertainty per year is reported in Appendix D.

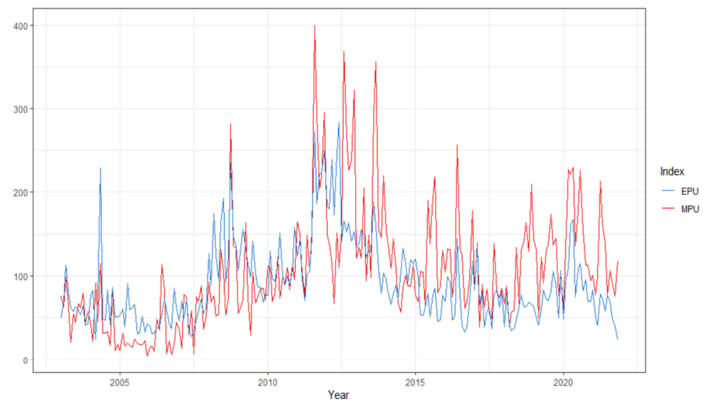
3.2 Comparison to Indian economic policy uncertainty

Since this is the first construction of a MPU index for India, the closest comparable index is the Indian EPU index developed by Baker et al. (2016), as displayed in Figure 2. A strong correlation of 0.61 confirms the relevant connection and expectations given the overlap in two of the three term categories. This can be also observed by the similar movements over the sample period. However, as emphasised above, MPU and EPU should be regarded as distinctive due to the subcategorical nature of the MPU. Hence, one can clearly observe differences, particularly in the volatility of the indices. Except for the spike in MPU after the win of the Congress party in the 2004 general elections, the peaks in MPU are generally lower than the EPU index until the Great Financial Crisis. Afterwards, volatility of the MPU index increased compared to EPU and remained higher until November 2021. In March and April 2016, the EPU and MPU index notably move in contrary directions. It is particularly interesting that the Eurozone Crises in 2012 had a more distinct effect on EPU. Although the Eurozone debt crisis was reported in the considered newspapers, particularly the elections in Greece and France in April and May 2012, it was not perceived as a bothersome topic of instability. Instead, during these months, MPU was more concerned with

a declining rupee and a pessimistic outlook on the balance of payments data. Nevertheless, shortly after the considerable decline in MPU, it peaked again following a troubling monsoon season and the downgrade of India's growth forecast. In contradistinction to uncertainty of monetary policy decision in the US, Eurozone politics did not appear to cause substantial uncertainty, which in turn would affect monetary policy in India. Subsequently, the Taper Tantrum in India and fear of US tapering of its QE policy had a much higher impact on MPU than on EPU.

Over the entire sample period, MPU follows a moderate counter-cyclical movement, similarly to the EPU index. Overall, it follows EPU closely until 2012. The differing approach in the categorisation and expanded set of terms could explain the differences between the two series. Furthermore, increased capital outflows in May and June 2013 or newly introduced reforms such as the inflation targeting in 2016 could further be adduced as an explanation for variation.

Figure 2: Comparison between MPU and EPU



This figure plots the Indian MPU compared to the Indian EPU index proposed by Baker et al. (2016) from January 2002 to November 2021 with data available at www.policyuncertainty.com.

4 Data

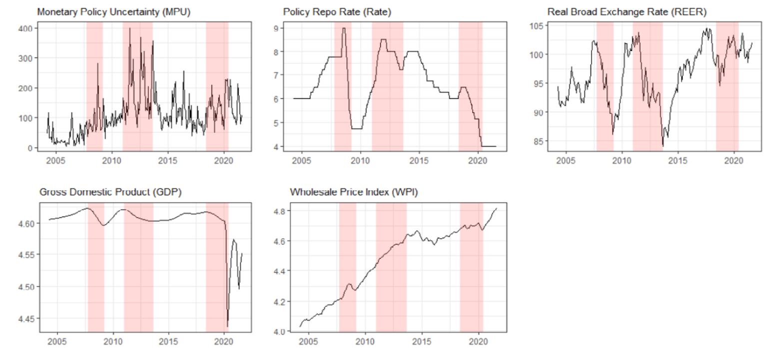
MPU in India is measured using the proposed index in section 3.1. In order to identify broader do mestic macroeconomic effects of MPU shocks, this paper takes into consideration the log-transformed Gross Domestic Product (GDP) (log(GDP)), while contemporaneously controlling for movements in the policy repo rate adopted by the RBI (Rate). In this sense, the RBI policy is used as a control instrument for Indian monetary policy (Caggiano, Castelnuovo, & Pellegrino, 2017). Fur ther, the log-transformed Broad Indian Rupee real exchange rate (log(REER)), which is calculated as weighted averages of bilateral exchange rates adjusted by relative consumer prices (BIS, 2021), and the log-transformed Wholesale Price Index (log(WPI)) are considered. The WPI consists of 435 commodities and was deliberately chosen instead of the Consumer Price Index, as the WPI represented India's official inflation rate until 2014 (T. Ghosh et al., 2021). Even after the switch to the Consumer Price Index (CPI), it was still reported and considered a stable inflation indicator. Furthermore, the CPI was firstly introduced in 2012 and can therefore not be taken as an inflation measure over the entire sample period. MPU, as well as macroeconomic data will be used at a monthly frequency. While most literature for MPU also considers growth rates of employment, the availability of employment data, especially on a monthly basis, for the time period has proven to be difficult. This is mainly due to the large unorganised sector of employment in India, which amounts to about 90% according to Ghate and Kletzer (2016).

Data for these variables is retrieved from the OECD Main Economic Indicators Databank (GDP), RBI Databank (Policy Rate), Office of Economic Adviser (WPI) and Board of Governors of the Federal Reserve System (Broad Real Exchange Rate). Due to the availability of data, the time period considered for the remainder of this paper includes April 2004 up to August 2021. The time series graphs of all variables used in this paper are also plotted in Figure 3.

Table 2: Descriptive Statistics For the Full Sample

	Obs.	Min.	Max.	Mean	SD	Skew.	Kurt.
Monetary Policy Uncertainty Data							
MPU	209	4.514	400.174	107.703	71.103	1.293	5.343
Monthly Macroeconomic Data							
log(GDP)	209	4.437	4.623	4.604	0.025	-3.784	19.002
log(WPI)	209	4.031	4.816	4.473	0.222	-0.542	1.866
Policy Repo Rate	209	84.110	104.590	96.571	4.757	-0.359	2.258
Real Exchange Rate	239	95.977	104.590	84.110	4.9074	-0.1268	1.9396

Figure 3: Time-series graphs of macroeconomic explanatory variables.



This figure shows the macroeconomic data for the 2004:04-2021:08 period after data transformation. Red shaded areas indicate OECD recession periods and were retrieved from https://www.oecd.org/sdd/leading-indicators/CLI-components-and-turning-points.pdf.

4.1 Construction of the index

To assure the existence of a statistical equilibrium, that is for the stochastic process to be ergodic, the previously introduced macroeconomic time series have been tested for stationarity. Controlling for stationarity and other variables is essential to prevent the rise of complications or spurious regressions in multivariate modelling, for instance, due to outliers or structural shifts. Stationarity is achieved if the first and second moments are time-invariant. As defined by Lütkepohl (2005), any stochastic process $y_t = y_1, y_2, \ldots, y_t$ in this case each variable MPU, ln(GDP), ln(WPI), Rate, and Reer, is stationary if it can fulfill two conditions:
Firstly,

$$E(y_t) = \mu \quad \text{for all} \quad t \in T,$$

which infers that each observation in the processes y_t has the same finite mean vector μ and therefore fluctuates around a constant mean.
Secondly,

$$E[(y_t - \mu_y)(y_{t-h} - \mu_y)] = y_h \quad \text{for all} \quad t \in T$$

and all integers h such that

$$t - h \in T.$$

This latter condition guarantees time-invariance and independence of the covariances on t. It is of importance to investigate the stationarity of time series, as non-stationarity can lead to spurious regressions, in which there are merely contemporaneous relationships of the variables instead of causal ones. Most often, the Augmented Dickey-Fuller (ADF) test or the Phillips-Perron (PP) test are utilised to test for a unit root. While the ADF test is primarily used as a standard in the literature, the PP test extends the unit root analysis, since it considers serial correlation, as well as heteroskedasticity in the error term. The ADF test equation includes the lagged changes of the variables as regressors and can be summarised as

$$\Delta y_t = c + \phi y_{t-1} + \sum_{j=1}^{p-1} \alpha_j \Delta y_{t-j} + u_t \tag{1}$$

The PP test extends this equation to include a centred time trend variable. For an integrated process, $\alpha(1) = 1 - \alpha_1 - ... - \alpha_p = 0$, both test the null hypothesis ($H_0 : \phi = 0$). The alternative hypothesis claims stationarity of the process, that is $\alpha(1) = 0$ or $H_1 : \phi < 0$. To interpret the results of the tests, the t-statistic of the coefficient ϕ can be regarded as significant. All variables of this analysis are tested on trend and intercept using the Akaike information criterion (AIC) with a maximum of 10 lags. The results are summarised in Table 3.

The presence of a unit root, that is non-stationarity, cannot be dismissed for any of them. To ensure weak stationarity, the variables were first-differenced and stationarity was confirmed.

Table 3: Stationarity Test for Macroeconomic Data

	Adj. t-Stat	1% crit. val.	5% crit. val.	10% crit. val.
Augmented Dickey- Fuller Test				
log(GDP)	-1.233	-4.005	-3.433	-3.140
log(WPI)	-1.303	-4.004	-3.432	-3.140
Policy Repo Rate	-2.171	-4.004	-3.432	-3.140
Real Exchange Rate	-2.861	-4.003	-3.432	-3.140
Phillips-Perron Test				
log(GDP)	-2.610	-4.003	-3.432	-3.139
log(WPI)	-1.487	-4.003	-3.432	-3.139
Policy Repo Rate	-1.843	-4.003	-3.432	-3.139
Real Exchange Rate	-2.619	-4.003	-3.432	-3.139

4.2 Cointegration

The variables in a time series vector yt are considered to be cointegrated of order (1, 1), $y_t \sim CI(1, 1)$, if all components of y_t are nonstationary with a unit root - that is if they are $I(1)$ - and there exists a linear combination $\beta' y_t$ is $I(0)$ (Hamilton, (2020). The nonzero (n x 1) vector β is called a cointegration matrix. Given the all non-zero scalar b, vector β is not unique, since $b\beta' \boldsymbol{y}_t$ is $I(0)$. In order for $\mathbf{B}'\mathbf{y}$ to be a stationary $(h \times 1)$ vector, there can be $h > n$ linear independent $(n \times 1)$ vectors $\beta_1, \beta_2, \ldots, \beta_h$. Then, $\mathbf{B}'$ can be expressed as the the following $(h \times n)$ matrix:

$$\mathbf{B}' = \begin{bmatrix} \beta_1' \\ \beta_2' \\ \vdots \\ \beta_h' \end{bmatrix}$$

Suppose such a relationship is assumed, it is desirable to test the variables for the presence of coin tegrating vectors. Thus, before analysing the dynamics of the baseline data, this test is conducted. To investigate the presence of any stochastic trends in the non-stationary level series, that is whether they are cointegrated, it is common in practice to use statistical tests to determine the cointegration rank r. Estimates can be obtained by maximising the log-likelihood for the given r. Therefore, the likelihood of different r is compared in the likelihood ratio test. The widely recognised Johansen Cointegration test (1991) is applied to analyse any long-run relationships between the data with the presence of r cointegrated vectors by evaluating the Trace test or the Maximum Eigenvalue test. Essentially, the Johansen test describes several steps that are required to estimate the maximum Gaussian log- likelihood of the model parameters (Neusser, (2016)). Firstly, one considers the variables y_t=[MPU$_t$, log(GDP$_t$), log(WPI$_t$), Rate$_t$, REER$_t$] in a VAR model of order p[2], that is,

$$y_t = c + \beta_1 y_t{-}1 + \beta_2 y_t{-}2 + \cdots + \beta_p y_t{-}p + u_t,$$

which can be written in error correction format (this step will be further explained in section 4) as

$$\Delta y_t = c + \Pi y_{t-1} + \Gamma 1 \Delta y_{t-1} + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t,$$

2 The choice of p was suggested by minimising AIC for a maximum of eight lags. The lag order selection can be found in Appendix 6.

where u_t are independent Gaussian errors with mean zero and variance Σ. To find the rank of matrix Π, Δy_t and y_{t-p} are regressed against c (the deterministic trend) and on the lagged values of Δy_{t-1} through Δy_{t-p+1}. As a next step, the residuals from these regressions are given in $(m \times 1)$ vectors of r_{0t} = residuals of regressing Δy on c, $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$ and r_{1t} = residuals of regressing $y-1$ on c, $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$ (Franses, Franses, et al., 1998). Thus, the product matrices are

$$\mathbf{\Sigma}_{00} := T^{-1}\sum_{i=1}^{T} r_{0t}r_{0t}'$$

$$\mathbf{\Sigma}_{11} := T^{-1}\sum_{i=1}^{T} r_{1t}r_{1t}'$$

$$\mathbf{\Sigma}_{01} := T^{-1}\sum_{i=1}^{T} r_{0t}r_{1t}'$$

$$\mathbf{\Sigma}_{10} := T^{-1}\sum_{i=1}^{T} r_{1t}r_{0t}'$$

Before obtaining the maximum of the Gaussian log-likelihood function for a given cointegration rank$(\Pi) = r_0$, the covariance matrix of the residuals $\hat{\Sigma}(\beta)$ has to be minimised over β. Then, λ_i becomes the solution to the eigenvalue problem $|\lambda\Sigma_{11} - \Sigma_{01}\Sigma_{00}^{-1}\Sigma_{10}| = 0,$ which leaves the eigenvectors that coincide to the r largest eigenvalues $\hat{\lambda}_1, \ldots \hat{\lambda}_r$. Finally, the log-likelihood function yields:

$$\ell = -\frac{Tn}{2}\ln\pi - \frac{Tn}{2} - \frac{T}{2}\ln\det\mathbf{\Sigma}_{00} - \frac{T}{2}\sum_{i=1}^{r}\ln(1-\lambda_i)$$

Further, the Maximum Eigenvalue and Trace test can be applied for determining the co-integrating rank h.

The Trace test statistics can be expressed as

$$\begin{aligned}\operatorname{Trace}(K, r_0) &= 2\left[\ln l(K) - \ln l(r_0)\right] \\ &= T\left[-\sum_{i=1}^{K}\ln(1-\lambda_i) + \sum_{i=1}^{r_0}\ln(1-\lambda_i)\right] \\ &= -T\sum_{i=r_0+1}^{K}\ln(1-\lambda_i)\end{aligned}$$

If the statistic of either test exceeds the critical values, the test will be rejected. The primary H_0 of h that is treated as the co-integration rank. To put this into perspective, it is essential to fully understand the values $\lambda 1, \ldots, \lambda_K$. Since they reflect the canonical correlation between a linear combination of levels, $\beta'_i y_t$, and any linear combination of differenced variables $\eta'_i \Delta y_{i,t}$, the magnitude of λ_i is an indication of how strongly the linear relation $\beta'_i y_t$ is correlated with the stationary process $\eta'_i \Delta y_{i,t}$. If there is indeed r_0 cointegration relations, then on the one hand $\beta'_1 y_t, \ldots, \beta'_{r0} y_t$ should be stationary, meaning that $\lambda_1, \ldots, \lambda_{r0}$ are large. On the other hand, if $\beta'_{r0+1} y_t, \ldots, \beta'_K y_t$ is not stationary, it concludes that $\lambda_{r0+1}, \ldots, \lambda_K$ are small. Therefore, Trace$(K, r0) = -T\Sigma^K_{i=r0+1} \ln(1-\lambda i)$ should be small under the null hypothesis rank$(\Pi) = r_0$. The maximum eigenvalue tests the significance of estimated eigenvalues themselves, or whether cointegration rank is $r-1$ or r, that is

$$H_0 : r = r_0 \text{ versus } H_1 : r = r_0 + 1$$

The maximum eigenvalue test statistic is given by

$$\lambda_{\max} = -T\ln(1-\lambda_{r0+1})$$

The final results testing the cointegrating rank of the baseline data are found in Table 4.2.

Table 4: Johansen Cointegration Test

Hypothesised No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Max-Eigenvalue Statistic	0.05 Critical Value
r = 0	0.165768	74.88202	69.81889	37.15507	33.87687
r ≤ 1	0.084163	37.72695	47.85613	18.02300	27.58434
r ≤ 2	0.064580	19.70396	29.79707	13.68583	21.13162
r ≤ 3	0.028829	6.018128	15.49471	5.996910	14.26460
r ≤ 4	0.000103	0.021218	3.841465	0.021218	3.841465

Following Lütkepohl, Saikonnen, and Trenkler (2001), the trace test will be regarded as superior to the Maximum Eigenvalue test and significant evidence is found to conclude that the system pos sesses the cointegration rank I(1). Also based on economic theory, cointegration is suggested. On the one hand, the Fisher effect implies a relationship between the inflation rate and interest rate. On the other hand, theory suggests a relationship between the real exchange rate and the interest, often referred to as interest rate parity.

5 Modelling the relationship between monetary policy uncertainty and economic variables

This section outlines the baseline VECM model and illustrates the test for autocorrelation. Section 5.2 explains the impulse response analysis and section 5.3 describes the estimation procedure to bootstrap the impulse responses.

5.1 VECM and VAR model specifications

The analysis of effects of MPU on economic growth is initiated by specifying the model of variables

$$y_t = [\text{MPU}_t, \log(\text{GDP}_t), \log(\text{WPI}_t), \text{Rate}_t, \text{REER}_t]$$

Except if indicated in any case, this is the variable ordering of the Cholesky decomposition. The ordering of the variables will be of importance for the analysis of impulse responses and follow decreasing order of exogeneity. That is, MPU is allowed to influence all following variables instan taneously, while GDP depends on a shock to MPU and its own shock contemporaneously. This follows the approach taken by Ghosh et al.(2021), Ueda (2010), as well as economic reasoning as GDP will react to changes in policy variables or other financial signals only with a lag. In line with Ueda (2010), inflation is also allowed to be affected by a shock in MPU and GDP, besides its own shock. While Baker et al. (2016) place the federal reserve funds rate before industrial production and employment, a different approach will be taken here. Non-policy variables such as output or inflation are often ordered Wold-casually[3] prior to the policy rate in monetary policy literature due to their slow-moving nature. Further, Friedman (1961), as well as Bernanke and Gertler (1995) provide evidence that monetary policy represented by the policy rate does not

3 The proposition is provided by the Wold Decomposition Theorem (1938), which states that every covariance stationary, non-deterministic stochastic time series x_t can be expressed as the sum of two uncorrelated processes, a deterministic component z_t (which could be assumed to be the mean term) and a component y_t that is a process with an infinite MA representation of $\sum_{i=0}^{\infty} \phi_i u_{t-i}$ (Neusser, 2016). Hence, $x_t = \sum_{i=0}^{\infty} \phi_i u_{t-1} + z_t$ where it is assumed that $\phi_0 = 1$ and $\sum_{i=0}^{\infty} \phi_j^2 < \infty$. The term u_t represents a Gaussian white noise process. The Wold Decomposition Theorem essentially proposes that every stationary, purely nondeterministic process can be approximated by a finite order VAR process (Lütkepohl, (2005)) .

contemporaneously affect inflation. The real exchange rate is ordered last, as Kim and Roubini (2000) demonstrate that the exchange rate only influences inflation with a lag and acts as a transmission channel of uncertainty shocks.

First introduced by Sims (1980), the methodology of Vector Autoregression (VAR), has become an essential tool for the analysis of dynamic relations between macroeconomic variables. For the set of four time series variables $y_t = \text{MPU}_t, \log(\text{GDP}_t), \log(\text{WPI}_t), \text{Rate}_t, \text{REER}_t$, a VAR model captures their dynamic interactions. In a system of weakly stationary processes, the variables are assumed to be jointly endogenous and each component of the vector y_t depends on its own lagged values and the lagged values of all other following variables (Kirchgässner et al., (2012)). In general, the VAR(p) model (VAR with p-th lagged order) has the form

$$y_t = c + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \tag{2}$$

where

$y_t = (y_{1,t}, \dots, y_{K,t})'$ (K time series variables)
$c = (c_1, \dots, c_K)'$ is a fixed vector of intercept
$ut = (u_{1,t}, \dots, u_{K,t})'$ is a vector of error

and A_i's coefficient matrices with $(K \times K)$ dimensions. The error term u_t is a white noise if it fulfills the condition to be zero-mean independent ($E(u_t) = 0$), possesses a time-invariant positive definite covariance matrix, $E(u_t u_t') = \Sigma_u$, and u_t and u_s to be independent for $s \neq t$ with finite fourth moments (Pfaff, 2008).

However, components of y_t can possess a linear combination or some long-run equilibrium if they are cointegrated. While VAR models are widely used in the identification of policy shocks, Engle and Granger (1987) pointed to the cointegration relationships among endogenous variables, which led to the development of the Vector Error Correction Model (VECM). As the previous cointegration analysis revealed, the variables of this analysis are cointegrated of order one. According to Lütkepohl et al. (2005) in the present case, it is appropriate to consider a VECM instead of a VAR model by subtracting y_{t-1} from both sides of equation 2 and rearranging to obtain the form,

$$y_t - y_{t-1} = -(I_K - A_1 - \dots - A_p) y_{t-1} - A_2 y_{t-1} + \dots + A_{p-1} y_{t-p+1} + u_t \tag{3}$$

or

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \tag{4}$$

where the parameter matrices are $\Pi = -(I_K - A_1 - \cdots - A_p)$, and $\Gamma_i = -(A_{i+1} + \cdots + A_p)$ for $i = 1, \ldots, p - 1$. In section 4.2, all the variables have been determined to be $I(1)$. In equation 4, y_{t-1} is the only term left that contains $I(1)$ variables. Consequently, for Δy_t to be $I(0)$, Πy_{t-1} must be $I(0)$ by assumption. If rank(Π) = r, then Π can be expressed as the product of $(K \times r)$ matrices α and β. Both matrices then have the same rank, such that rank(α) = rank(β) = r. Equation 4 can therefore also be rewritten as

$$\Delta y_t = \alpha\beta y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \tag{5}$$

As a consequence, $\beta' y_{t-1}$ is a stationary process because Π becomes $I(0)$ through the premultiplication of an $I(0)$ vector α. Therefore, $\beta' y_{t-1}$ includes the cointegrating relations and yt contains r = rank(Π) linearly independent cointegrating relations (Lütkepohl, 2005). The differing compo nent to the level VAR representation is given by the error correction term (ECT)= $\beta' y_{t-1}$, which enables y to be pushed back towards the long-run equilibrium. Πy_{t-1} can also be expressed as $\Pi y_{t-1} = \alpha$ECT, with α often referred to as the loading matrix that contains the weights of the cointegrating relations in each equation. Only the product of α and β, $\alpha\beta'$, is estimated individu ally, as there are many different possible matrices for both components.

Considering that most macroeconomic level series are constantly developing, level series do not allow for interpretation and comparison over time. Especially for the comparison of economic ef fects over a long time span and across nations, growth rates are advantageous. When applying a VECM model, it is important to note that the macroeconomic variables are first-differenced, which also brings about stationarity and allows for interpretation as growth rates of the original series. The baseline data was first-differenced beforehand and tested on stationarity, which revealed stationarity after first-differencing for all variables. Thus, in the VECM model, the variables are not first-differenced again. It could certainly be argued that by first differencing, the relationship between the original variables, as shown by the cointegration, may be lost. However, particularly the log-transformed series can now be investigated in their growth rates, which is interesting by its own merit, as mentioned before. Resultingly, with the application of the VECM model, this paper will refrain from using level series and instead investigate the effects on economic growth indicators for the remainder of this paper.

It is essential to mention that there is a one-to-one corresponding of the VECM and VAR levels representation (equation 2). The VECM was directly obtained by subtracting y_{t-1} from both sides of equation 2. This also implies that all methods to determine the lag order (FPE, BIC, AIC, and SC) in stationary and nonstationary VAR are applicable in the VECM. If the lag order is chosen so that p = p* in VAR, then the order of lagged difference in the VECM is p* − 1. For the underlying VAR, the AIC criteria suggests three lags, hence for the baseline specification, a VECM(2) is estimated. Generally, a VAR process is assumed to be stable if the determinant of the autoregressive operator does not have any root in or on the complex unit circle (Lütkepohl & Krätzig, 2004). That is,

$$\det(I_n - A_1 z - \cdots - A_n z^p) = 0 \text{ for } |z| \leq 1,$$

where I_n is the identity matrix[4]. Since the baseline VECM has five endogenous variables and one cointegrating vector, the lagpolynomial has four eigenvalues with unit moduli and further moduli not excessively near one, which can be regarded as evidence for the stability of the system according to Kilian and Lütkepohl (2017). To obtain confidence in the built model, the Lagrange-multiplier test was also conducted. Table 5 presents the test for serial correlation on the VECM model resid uals in equation 4. The null hypothesis states that no autocorrelation is present and it cannot be rejected at lag two.

Table 5: Lagrange-multiplier Test

Lag	chi2	df	Prob > chi2
1	38.57416	25	0.0406
2	33.68923	25	0.1147
3	37.65629	25	0.0500

H0: no autocorrelation at lag order

5.2 Impulse response analysis

Any relations between the various variables in the VECM model are difficult to grasp by merely analysing the coefficients. Therefore, to infer information from the marginal response of an endogenous variable of the system to an impulse of another endogenous variable, IRFs are mostly considered. In the case of orthogonal components and no cointegration of the process, shocks to one variable can be analysed in terms of a Mov

4 It constitutes a matrix of n n × n dimensions zeros everywhere except for the main diagonal, which is filled with ones.

ing Average (MA) representation

$$y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \cdots + \Phi_s u_{t-s} \quad (6)$$

where $\Phi_0 = I_K$ and the $\Phi_s = s_j$=1 $\Phi_{s-j}A_j$, $s = 1, 2, \ldots$ (Lütkepohl, 2005). The resulting coefficients are the responses upon an impulse hitting the system. Within the matrices Φ_s, the (i, j)th elements determine the expected responses of each variable in the system (y_i,$t+s$) to a one-time impulse in MPU (y_{jt}). To measure the change in the responding variables of Y_t, one refers to u_{it}. The elements of Φ_s thus include the impulse responses of MPU, ln(GDP), ln(WPI), Rate, and Reer (y_t) following an impulse of y_{jt}.

The Wold MA representation applies primarily to impulse responses of stationary and stable VAR(p) processes (Lütkepohl, 2005). However, in this case, the cointegrated nature of variables has to be taken into consideration. Furthermore, the non-diagonal covariance matrix Σ_u also shows that resid uals can often be instantaneously correlated. To avoid indirect effects, it is necessary to transform the residuals into uncorrelated shocks with the help of suitable transformations. Hence, orthog onalisation is necessary to analyse the impact of a single, isolated shock. In order to obtain an orthogonal impulse response to a one-time shock, the Cholesky decomposition of the covariance matrix Σ_u is utilised. By orthogonalising the shocks so that $\varepsilon_t = B^{-1}u_t$, the covariance matrix is defined by $\Sigma_u = BB'$, and B being a lower triangular matrix (Lütkepohl, 2005). Then 6 can be re-written as

$$y_t = \Psi_0\varepsilon_t + \Psi_1\varepsilon_{t-1} + \cdots + \Psi_i\varepsilon_{t-i}, \quad (7)$$

or

$$y_t = \Sigma^{\infty}_{i=0}\Psi_i\varepsilon_{t-i} \quad (8)$$

where $\Psi_i = \Phi_i B$, (i = 0,1,2,...). In this case, $\Phi_0 = B$ is lower triangular because $\Phi_0 = I_n$. If a shock hits the first variable in the system, it has an instantaneous effect on all the following variables in the Cholesky ordering. The second variable then has an instantaneous effect on the variables except for y1t and so forth. It becomes clear that the ordering of the variables is crucial since this mechanism suggests a Wold causal chain of ordering, for which one variable can only affect the ones ordered afterwards.

Finally, the IRFs for each variable i in the system Y, that is MPU, ln(GDP), ln(WPI), Rate, and Reer, can be computed as it is following an impulse in MPU,

$$\mathrm{IRF}_{i,MPU} = \frac{\delta y_{i,t+s}}{\delta\varepsilon_{MPU,t}} = \frac{\delta y_{i,t}}{\delta\varepsilon_{MPU,t-s}} = \Psi_s \quad (9)$$

This computation of Ψs and Φs is the same as for stationary systems. Yet, the crucial difference is that for cointegrated systems, the effect of an impulse may not return to zero as the time goes to infinity, unlike in stable processes.

5.3 Bootstrapping of the Impulse Response Functions

The estimators of IRFs are obtained as, $\hat{\psi}_{ij,h} = \psi_{ij,h}\left(\hat{\alpha}, \hat{\beta}, \hat{\Gamma}_0, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_{p-1}\right)$ where $\hat{\alpha},\ \hat{\beta},\ \hat{\Gamma}_0, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_{p-1}$ are the estimated VECM parameter matrices as specified in equation 4. To display IRFs, it seems to be the standard in literate to visualise them by plotting 68%[5], 90%[6], or 95% confi dence intervals. Under the assumption that the resulting impulse responses have asymptotic normal distributions, they can be used to derive the confidence intervals (Benkwitz et al., 2001). However, the analytical determination of such intervals in VAR or VECM models only yields asymptotic statements and they are only valid statements with large sample sizes. Bootstrapping or simulation methods offer alternatives that generate a set of impulse-response functions with the help of Monte Carlo simulations, from which confidence bands are estimated. Thus, they are generally preferred due to their reliable sample inference. Based on the existing data set, a new data set is generated and the distribution of the estimated parameters can be determined. In the following, the bootstrap method as proposed by Benkwitz and Lütkepohl (2001) is applied:

1. Estimation of the model parameters

The VECM model parameters were estimated as put forward in section 5. With this, the sets of residuals ut are obtained.

5 Consider for example Husted et al. (2020), Trung (2019), Caggiano et al. (2017), Fontaine et al. (2017).

6 For instance Priyaranjan and Pratap (2020) use these intervals.

2. Generation of bootstrap residuals

To create the bootstrap residuals, Benkwitz and Lütkepohl (2001) propose to generate new residuals through a random generator and sample them by replacement, such that they could be reused indefinitely. Then the newly drawn residuals are re-centred by subtracting the mean of the residuals from each residual, that is $\hat{u}_T - \bar{u}$. In total, 600 sets of bootstrap residuals were created, so that the (600 × r * c) residual matrix, with r * c being rows and columns of the orthogonalised variance covariance, could be obtained. The covariance matrix was estimated like in standard asymptotic theory. Hence the residual matrix could be expressed as

$$\mathbf{u}_{i,t} = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,r*c} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,r*c} \\ \vdots & & & \\ u_{600,1} & u_{600,2} & \cdots & u_{600,r*c} \end{bmatrix}$$

3. Computing the bootstrap time series

With the help of the newly created residuals and the estimated parameters of the VECM model, it is possible to recursively generate a new bootstrap dataset and set $y^*_{-p+1}, \ldots, y^*_0 = (y_{-p+1}, \ldots, y_0)$. This is done using the levels representation in 2. In combination with the bootstrap residuals, the eigenvectors, variables in levels, regressors, and coefficients are used so that the 5-dimensional VAR(3) is such that

$$y_{1,t} = c_1 + a_{11}y_{1,t-1} + \ldots + a_{15}y_{5,t-3} + u_1$$

$$\vdots$$

$$y_{5,t} = c_5 + a_{51}y_{1,t-3} + \ldots + a_{55}y_{5,t-3} + u_5$$

For this VAR(3) model, Y is defined as a (K × T) matrix for K-variables over T sample period. Hence, it is defined $Y = (y_1, \ldots, y_t)_{5\times209}$, $A = (c, A_1, \ldots, A_3)_{5\times(Kp+1)}$, $U = (u_1, \ldots, u_{209})_{5\times209}$,

$$Z = \begin{pmatrix} 1 & \cdots & 1 \\ y_0 & \cdots & y_{209-1} \\ \vdots & & \vdots \\ y_{-3+1} & \cdots & y_{209-p} \end{pmatrix}_{(Kp+1)\times 209} .$$

Which can also be expressed as the reduced matrix form:

$$Y = AZ + U^* \tag{10}$$

4. Estimation of the IRF for each Bootstrap Dataset

To estimate the new VECM parameters $\Gamma_0, \Gamma_1, \ldots \Gamma_{p-1}$, α, β in equation 4 of the impulse response function $\hat{\psi}_{ij,h}$ again, the same cointegration rank r(1) is used. Benkwitz and Lütekepohl (2001) also explore the possibility to reestimate the cointegration matrix for each bootstrap sample. However, this approach was not taken and the cointegration matrix is not re-estimated. In the case of a levels VAR, $\hat{\phi}_{ij,h}$ could be estimated from the model's Wold MA representation.

5. Calculating the Bootstrap Confidence intervals

As the last step, confidence intervals are set up on the estimations. Benkwitz and Lütkepohl (Benkwitz et al., 2001) compute the confidence intervals - in this case 95% - as the $\gamma/2$ and $(1-\gamma)/2$ quantiles of the bootstrap parameter distribution

$$\mathcal{L}\left(\hat{\Psi}_t \mid \mathbf{y}_{-p+1}, \ldots, \mathbf{y}_0, \ldots, y_t\right).$$

6 Results

This section presents the results obtained from the impulse responses following a positive one standard deviation of MPU throughout 15 months.

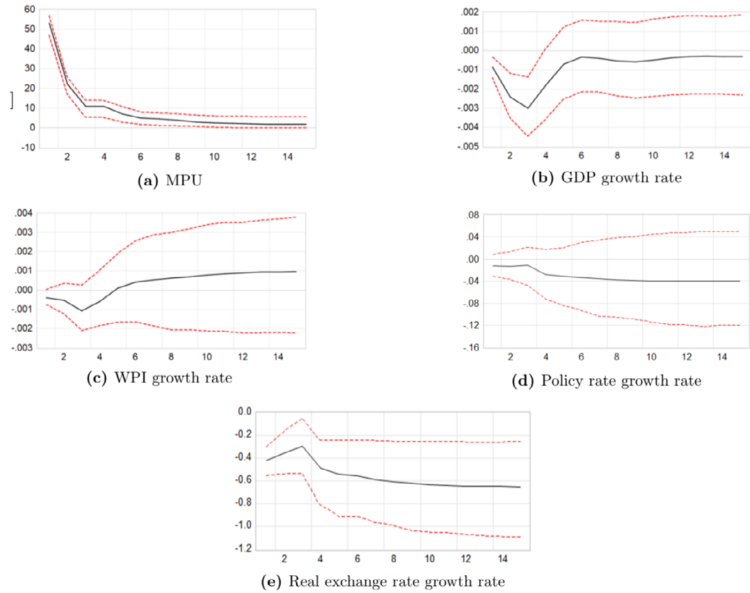
To test the robustness of the results, the VECM model is re-estimated with alternative specifications and variables in section 6.2.

6.1 Impulse Responses

The impulse responses are plotted in Figure 4 and summarise one standard deviation shock to MPU throughout the following 15 months. The displayed IRFs indicate that a one-standard-deviation MPU shock indeed has negative effects on all considered variables. The responses of the variables do not diminish over time, which is due to the nature of VECM models as mentioned in section 5. While Baker et al. (Baker et al., 2016) inspect Industrial Production in their estimation, GDP follows a similar pattern in the case of MPU. GDP reaches a significant negative peak after three months of 0.3%, recovering fairly quickly afterwards. The result is in line with Ghosh et al. (2021) and Bhagat et al. (2016), who both estimate a negative output effect from an EPU shock. The magnitude of the drop in GDP is similar to the response of Chinese output when confronted by an MPU shock, as evaluated by Huang and Luk (2020). Furthermore, a positive MPU shock mini mally decreases inflation upon impact by 0.05% before returning to normal levels after five months. Though, this response has to be considered with caution, as the result is not statistically signifi cant. Bearing in mind that MPU is a categorical uncertainty measure, it seems to be the case that MPU does not affect long-run fluctuations in inflation, which may be rather affected by broader fundamental economic disruptions. The response of the policy rate is negative for fifteen months following the shock. To lower the policy rate after elevated uncertainty would be in line with con ventional monetary policies of central banks to restore economic growth. Lastly, the real exchange rate shows an immediate significant depreciation with a permanent effect. Thus, an increase in MPU reduces output and inflation rate, while at the same time the real exchange rate experiences a great depreciation. For the long-lasting effects, an analysis of the overall influence of each of the macroeconomic variables in the system is investigated with the variance decomposition. Appendix B presents the variance decomposition for the VECM model over the sample period 2004-2021. It allows for analysing how much of the change in one variable is caused by an exogenous shock in another variable. The investigations show that MPU is not a significant determinant of inflation in the long run, however, it is a determinant for the real exchange rate. This is in line with the recent policy objectives of the RBI that take the stability of the exchange rate into close consideration. Over the long run, output is largely determined by itself, yet MPU

does have some influence, par ticularly in the short run. The statistically insignificant results for the inflation growth rate are in line with expectations since the variable decomposition reveals that inflation is largely driven by itself.

Figure 4: Impulse Response Functions



This figure displays the responses of the GDP and WPI growth rates to a one-standard-deviation shock in Indian MPU for the following 15 months. It also shows the responses for the policy interest rate, broad real exchange rate, and MPU. The parameters were estimated using a VCE model and bootstrapping. The model is estimated on monthly data from March 2004 to August 2021. The shocks were constructed with the use of a Cholesky decomposition with the ordering of the variables as follows: $y_t = [MPU_t, log (GDP_t), log (W PI_t), Rate_t, REER_t]$. The red dashed lines indicate the 95% confidence intervals.

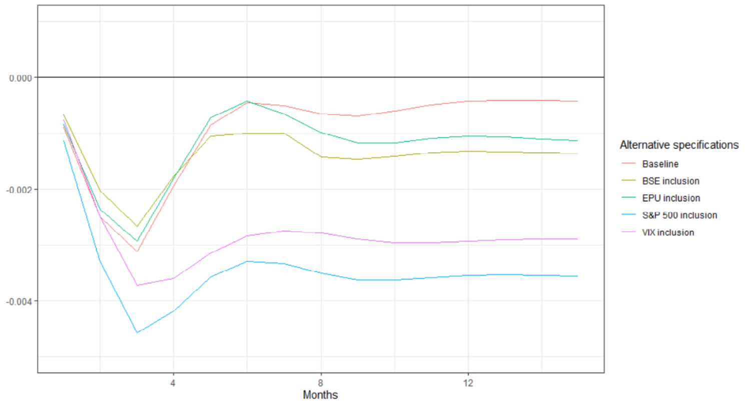
6.2 Robustness check

In order to consider the robustness of the baseline results, the VECM model is estimated with alternative specifications and additional variables. For this, the Indian EPU index[7] is included as an additional element of uncertainty. Following the approach by Baker et al. (2016) with their inclusion of the S%P 500 for the US, the Bombay Stock Exchange (BSE) is included in the estimation as a controlling variable for the In

7 Note that this index is available online http://www.policyuncertainty.com/ and updated regularly.

dian financial market. This is done due to the forward-looking nature of stock markets. As such, critical information may already be anticipated by the stock market. Further, also the US stock market S&P 500 and Volatility Index (VIX) are included as alternative specifications. India's close economic relationship with the US could suggest spillover effects of uncertainty. These effects are confirmed by Trung (2019) in the decline in India's price index following a US EPU shock to India. In the Cholesky decomposition, VIX and EPU are ordered after MPU, while the stock market indices BSE and S&P 500 are ordered last. Figure 6 shows the result of GDP growth rate following an MPU shock with the alternative specifications and the baseline estimate. While the baseline estimation is robust to the inclusions of more domestic uncertainty variables, the effect of an MPU shock with the consideration of the US stock exchange and volatility index seems to amplify the negative effect even more. Notably, the negative effect is more persistent in the latter two specifications and could suggest considerable spillover effects and linkages to the US economy.

Figure 6: GDP Growth Response to an MPU Shock with Alternative Specifications



This figure displays the response of the GDP growth rate to a one-standard-deviation following an Indian MPU shock using the VECM specification and identification as in Figure 4 on the same sample period. VIX and EPU are ordered after MPU in the respective specification. Confidence intervals are omitted for illustrative purposes.

7 Conclusion and discussion

This paper attempts to examine to what extent shocks of MPU affect economic indicators of the Indian economy. Constructing a new index for Indian MPU enables an investigation of economic effects following increased uncertainty in the realm of monetary policy. Especially, using monthly economic time series from 2004 to 2021, effects on GDP growth, inflation rate, real exchange rate, and policy rate following one standard deviation of MPU are measured by applying a VECM model to the economic assets. The IRFs are identified by making use of bootstrapping.

Summarising the results, this study finds significant and economically relevant effects of MPU on the Indian economy. The analysis points to cointegration between the macroeconomic level series and MPU. Further, it confirms results brought forward by the academic literature on MPU and EPU effects in India. Upon impact, all considered variables have a negative impact. GDP growth falls sharply by 0.3% before recovering to normal levels after 5 months, whereas the real exchange rate experiences a permanent depreciation. Inflation decreases in response to an increase in MPU at first, however, increases shortly afterwards. The response to inflation has to be considered with caution though, as the result is not significant over the 15 months. This does not come as a surprise, since the variable decomposition reveals that inflation is largely driven by itself. Bearing in mind that MPU is a categorical uncertainty measure, it seems to be the case that MPU does not affect long-run fluctuations in inflation, which may be rather affected by broader fundamental economic disruptions. Over the long run, output is largely determined by itself, yet MPU does have some influence, particularly in the short run.

The findings of the paper have several interesting implications not only for individual and institu tional investors but also for policymakers. The implications drawn from this paper can provide a powerful tool for understanding and acknowledging the potential of India as a significant emerging market and the impact of MPU on the Indian domestic economy and its growing importance in the international markets.

Despite cautious consideration, many unanswered questions and stimulating avenues last for further research. Not only can the construction of the Indian MPU be altered by changing the search terms and including even terms in Hindi, but the scope of newspapers could certainly also be widened to include mainland Hindi newspapers as well. For instance, Huang and Luk (2020) measure Chinese MPU with the inclusion of mainland China newspapers, which is more likely to capture the full effect of MPU. Furthermore, the construction of an MPU index is computationally intense and it would be exceptionally tedious to continue developing uncertainty series in this manner over longer samples. Identification of MPU through newspaper reports is a promising method of capturing uncertainty, however, in the future, this alone will not suffice adequately to recognise the causal impacts of monetary policy volatility. Hence, an approach brought forward by Stock and Watson (2018) could be adduced as an alternative. Essentially endogenous dynamic development of variables and uncertainty can be assessed with the utilisation of the persistent EPU index. While the origin of exogenous shocks is usually identified through internal instruments, such as restrictions (in this case the Cholesky decomposition), external instruments could be used, as proposed by Stock and Watson (2018). In this method, exogenous variables are used for the identification of effects on the sub-sample. This would enable to loosen restraints on the ordering of the variables and allow identification of various contemporaneous relationships. This leads to another key limitation encountered in this paper, which regards the utilised model. With the usage of a structural VECM model instead of a VECM, restrictions for structural shocks on the reduced-form could be imposed, which would allow the identification of the system of equations. For instance, defining one structural equation per variable presents one of them, which could be further relaxed and altered in follow up research. As a result of the chosen method in this paper, the magnitude of certain results may vary, thus it must be noted that over-or underestimation of uncertainty in the economy are a common phenomenon in the literature. This can, on the one hand, be caused by the inability to differentiate between MPU or EPU and other shocks that impact the economy. On the other hand, with the use of a newspaper index, the exact timing of shocks is sensitive to perception, thus effects may also be given too little weight. To alleviate this, differentiation between MPU shocks could be considered by introducing non-linearity into the model. This could potentially differentiate between smaller shocks and large shocks, like the Great Financial Crisis. As this paper found evidence of cointegration relations, any long-term relationships between Indian MPU and other economic variables such as industrial production growth estimates or private consumption growth could be investigated. Following this line of thought, transmission channels of MPU could be analysed more closely. Furthermore, even though the VAR model for spillover effects on other stock markets provides promising results, a closer investigation into the bilateral effects opens the need for further research. Finally, modelling business cycles distinctly with a Markov-switching VAR model or a

Threshold VAR model could also potentially provide opportunities for the extension of this paper.

8 References

Aleem, A. (2010). Transmission mechanism of monetary policy in india. *Journal of Asian Economics*, 21 (2), 186–197. doi: https://doi.org/10.1016/j.asieco.2009.10.001

Al-Mashat, R. (2003). Monetary policy transmission in india: Selected issues and statistical appendix. *Country report*, 3 , 261.

Anand, R., & Tulin, M. V. (2014). Disentangling india's investment slowdown. International Monetary Fund.

Arbatli, E. C., Davis, S. J., Ito, A., & Miake, N. (2017, May). *Policy uncertainty in japan* (Working Paper No. 23411). National Bureau of Economic Research. doi: https://doi.org/10.3386/w23411

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131 (4), 1593–1636. doi: https://doi.org/10.1093/qje/qjw024

Banerjee, S., Mohanty, M., et al. (2021). *Us monetary policy and the financial channel of the exchange rate: evidence from india*. Bank for International Settlements, Monetary and Economic Department.

Benkwitz, A., Lütkepohl, H., & Wolters, J. (2001). Comparison of bootstrap confidence intervals for impulse responses of german monetary systems. *Macroeconomic dynamics*, 5 (1), 81–100. doi: https://doi.org/10.1017/S1365100501018041

Bernanke, B. S. (1983). Irreversibility, uncertainty, and cyclical investment. *The quarterly journal of economics*, 98 (1), 85–106. doi: https://doi.org/10.2307/1885568

Bernanke, B. S., & Gertler, M. (1995). Inside the black box: the credit channel of monetary policy transmission. *Journal of Economic perspectives*, 9 (4), 27–48. doi: 10.1257/jep.9.4.27

Bhagat, S., Ghosh, P., & Rangan, S. (2016). Economic policy uncertainty and growth in india. *Economic and Political Weekly*, 72–81. BIS. (2021). About foreign exchange statistics. Retrieved 06.04.2022, from https:// www.bis.org/ statistics/about fx stats.htm

Bloom, N. (2009). The impact of uncertainty shocks. *econometrica*, 77 (3), 623–685. doi: https:// doi.org/10.3982/ECTA6248

Bloom, N. (2014). Fluctuations in uncertainty. *Journal of Economic Perspectives*, 28 (2), 153–76. doi: 10.1257/jep.28.2.153

Bloom, N., Bond, S., & Van Reenen, J. (2007). Uncertainty and investment dynamics. *The review of economic studies*, 74 (2), 391–415. doi: https://doi.org/10.1111/j.1467-937X.2007.00426.x

Caggiano, G., Castelnuovo, E., & Figueres, J. M. (2017). Economic policy uncertainty and unem ployment in the united states: A nonlinear approach. *Economics Letters*, 151 , 31–34. doi: https://doi.org/10.1016/j.econlet.2016.12.002

Caggiano, G., Castelnuovo, E., & Pellegrino, G. (2017). Estimating the real effects of uncertainty shocks at the zero lower bound. *European Economic Review*, 100 , 257–272. doi: https:// doi.org/10.1016/j.euroecorev.2017.08.008

Cevik, S., & Erduman, Y. (2020). Measuring monetary policy uncertainty and its effects on the economy: The case of turkey. *Eastern European Economics*, 58 (5), 436–454. doi: https:// doi.org/10.1080/00128775.2020.1798161

Chen, H., & Tillmann, P. (2021). Monetary policy uncertainty in china. Journal of International Money and Finance, 110 , 102309. doi: https:// doi.org/10.1016/j.jimonfin.2020.102309 Chudik, A., & Fratzscher, M. (2011). Identifying the global transmission of the 2007–2009 financial crisis in a gvar model. *European Economic Review*, 55 (3), 325–339. doi: https://doi.org/ 10.1016/j.euroecorev.2010.12.003

C¸ olak, G., Durnev, A., & Qian, Y. (2017). Political uncertainty and ipo activity: Evidence from us gubernatorial elections. *Journal of Financial and Quantitative Analysis*, 52 (6), 2523–2564. doi: https://doi.org/10.1017/S0022109017000862

Colombo, V. (2013). Economic policy uncertainty in the us: Does it matter for the euro area? Economics Letters, 121 (1), 39–42. doi: https://doi.org/10.1016/j.econlet.2013.06.024 Constantinescu, C., Mattoo, A., & Ruta, M. (2017). Trade developments in 2016. Dua, P. (2020). Monetary policy framework in india. *Indian Economic Review*, 55 (1), 117–154. doi: https://doi.org/10.1007/s41775-020-00085-3

Eberly, J. C. (1994). Adjustment of consumers' durables stocks: Evidence from automobile pur chases. *Journal of political Economy*, 102 (3), 403–436. doi: https://doi.org/10.1086/261940

Eichengreen, B. (2013). *International policy coordination: the long view. In Globalization in an age of crisis: Multilateral economic cooperation in the twenty-first century* (pp. 43–82). University of Chicago Press.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. Econometrica: journal of the Econometric Society, 251–276. doi: https://doi.org/10.2307/1913236 Fern´andez-Villaverde, J., Guerr´on-Quintana, P., Kuester, K., & Rubio-Ram´ırez, J. (2015). Fiscal volatility shocks and economic activity. *American Economic Review*, 105 (11), 3352–84. doi: 10.1257/aer.20121236

Fontaine, I., Didier, L., & Razafindravaosolonirina, J. (2017). Foreign policy uncertainty shocks and us macroeconomic activity: Evidence from china. *Economics Letters*, 155 , 121–125. doi: https://doi.org/10.1016/j.econlet.2017.03.034

Franses, P. H., Franses, P. H. B., et al. (1998). *Time series models for business and economic forecasting*. Cambridge university press.

Fratzscher, M., Lo Duca, M., & Straub, R. (2012). A global monetary tsunami? on the spillovers of us quantitative easing. *On the Spillovers of US Quantitative Easing* (October 19, 2012). doi: https://dx.doi.org/10.2139/ssrn.2164261

Friedman, M. (1961). The lag in effect of monetary policy. *Journal of Political Economy*, 69 (5), 447–466. doi: https://doi.org/10.1086/258537

Gabauer, D., & Gupta, R. (2018). On the transmission mechanism of country-specific and interna tional economic uncertainty spillovers: Evidence from a tvp-var connectedness decomposition approach. *Economics Letters*, 171 , 63–71. doi: https://doi.org/10.1016/j.econlet.2018.07.007

Ghate, C., & Kletzer, K. M. (2016). *Monetary policy in india: A modern macroeconomic perspective*. Springer.

Ghosh, A. R., Qureshi, M. S., & Jang, E. S. (2016). Capital flows and capital controls in india: confronting the challenges. In *Monetary policy in india* (pp. 299–333). Springer. doi: https:// doi.org/10.1007/978-81-322-2840-0 10

Ghosh, T., Sahu, S., & Chattopadhyay, S. (2021). Inflation expectations of households in india: Role of oil prices, economic policy uncertainty, and spillover of global financial uncertainty. *Bulletin of Economic Research*, 73 (2), 230–251. doi: https://doi.org/10.1111/boer.12244

Greenspan, A. (2003). *Monetary policy under uncertainty*. Retrieved 10.05.2022, from https:// www.federalreserve.gov/boarddocs/Speeches/2003/20030829/default.htm#pagetop Hamilton, J. D. (2020). Time series analysis. Princeton university press.

Hammond, G., Kanbur, S. R., & Prasad, E. (2009). *Monetary policy frameworks for emerging markets*. Edward Elgar Publishing.

Hardouvelis, G. A., Karalas, G., Karanastasis, D., & Samartzis, P. (2018). Economic policy uncertainty, political uncertainty and the greek economic crisis. Political Uncertainty and the Greek Economic Crisis (April 3, 2018). doi: https://dx.doi.org/10.2139/ssrn.3155172 Huang, Y., & Luk, P. (2020). Measuring economic policy uncertainty in china. *China Economic Review*, 59 , 101367. doi: https://doi.org/10.1016/j.chieco.2019.101367

Husted, L., Rogers, J., & Sun, B. (2020). Monetary policy uncertainty. Journal of Monetary Economics, 115 , 20–36. doi: https://doi.org/10.1016/j.jmoneco.2019.07.009 Iyke, B. N. (2020). Economic policy uncertainty in times of covid-19 pandemic. *Asian Economics Letters*, 1 (2), 17665. doi: https://doi.org/10.46557/001c.17665

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. Econometrica: *journal of the Econometric Society*, 1551–1580. doi: https://doi.org/10.2307/2938278

Julio, B., & Yook, Y. (2012). Political uncertainty and corporate investment cycles. The Journal of Finance, 67 (1), 45–83. doi: https://doi.org/10.1111/j.1540-6261.2011.01707.x Julio, B., & Yook, Y. (2016). Policy uncertainty, irreversibility, and cross-border flows of capital. *Journal of International Economics*, 103 , 13–26. doi: https://doi.org/10.1016/j.jinteco.2016 .08.004

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105 (3), 1177–1216. doi: https://doi.org/10.1257/aer.20131193

Kelly, B., P´astor, L., & Veronesi, P. (2016). The price of political uncertainty: Theory and evidence from the option market. *The Journal of Finance*, 71 (5), 2417–2480. doi: https://doi.org/ 10.1111/jofi.12406

Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

Kim, S., & Roubini, N. (2000). Exchange rate anomalies in the industrial countries: A solution with a structural var approach. *Journal of Monetary economics*, 45 (3), 561–586. doi: https:// doi.org/10.1016/S0304-3932(00)00010-6

Kirchgässner, G., Wolters, J., & Hassler, U. (2012). I*ntroduction to modern time series analysis*. Springer Science & Business Media.

Leduc, S., & Liu, Z. (2016). Uncertainty shocks are aggregate demand shocks. *Journal of Monetary Economics*, 82 , 20–35. doi: https://doi.org/10.1016/j.jmoneco.2016.07.002

Li, X.-l., Balcilar, M., Gupta, R., & Chang, T. (2016). The causal relationship between economic policy uncertainty and stock returns in china and india: Evidence from a bootstrap rolling window approach. *Emerging Markets Finance and Trade*, 52 (3), 674–689. doi: https:// doi.org/10.1080/1540496X.2014.998564

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Lütkepohl, H., & Kr¨atzig, M. (2004). *Applied time series econometrics*. Cambridge university press.

Lüutkepohl, H., Saikkonen, P., & Trenkler, C. (2001). Maximum eigenvalue versus trace tests for the cointegrating rank of a var process. *The Econometrics Journal*, 4 (2), 287–310. doi: https://doi.org/10.1111/1368-423X.00068

Mohan, R. (2006a). Evolution of central banking in india. *Reserve Bank of India Bulletin, June, 17*.

Mohan, R. (2006b). Monetary policy and exchange rate frameworks: The indian experience. *Reserve Bank of India Bulletin, June*.

Neusser, K. (2016). *Time series econometrics*. Springer.

P´astor, L., & Veronesi, P. (2013). Political uncertainty and risk premia. *Journal of financial Economics*, 110 (3), 520–545. doi: https://doi.org/10.1016/j.jfineco.2013.08.007

Pesce, M. A., et al. (2008). Transmission mechanisms for monetary policy in emerging market economies: what is new? In *Participants in the meeting* (p. 131).

Pfaff, B. (2008). *Analysis of integrated and cointegrated time series with r*. Springer Science & Business Media.

Popp, A., & Zhang, F. (2016). The macroeconomic effects of uncertainty shocks: The role of the financial channel. *Journal of Economic Dynamics and Control*, 69 , 319–349. doi: https:// doi.org/10.1016/j.jedc.2016.05.021

Priyaranjan, N., & Pratap, B. (2020). Macroeconomic effects of uncertainty: a big data analysis for india. *Priyaranjan, N., & Pratap, B.(2020). Macroeconomic Effects of Uncertainty: A Big Data Analysis for India*. RBI Working Paper (04). doi: https://doi.org/10.1016/j.asieco.2009 .10.001

RBI. (1934). *Reserve bank of india (rbi) act, 1934*. Retrieved 03.05.2022, from https://www.rbi.org.in/Scripts/OccasionalPublications.aspx?-head=Reserve% 20Bank%20of%20India%20Act

Rebello, J. (2016). *Raghuram rajan pours cold water on speculations, says won't accept second term*. Retrieved 12.05.2022, from https:// economictimes.indiatimes.com/news/economy/policy/raghuram-rajan-pours-cold-water-on-speculations-says-wont -accept-second-term/articleshow/52809474.cms

Reddy, Y. V. (2010). Uneasy on cloud nine. *Hindustan Times*. Retrieved 27.04.2022, from https://bit.ly/37S6azq

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1–48. doi: https://doi.org/10.2307/1912017

Singh, K., & Kalirajan, K. (2007). Monetary transmission in post-reform india: an evalua tion. *Journal of the Asia Pacific Economy*, 12 (2), 158–187. doi: https://doi.org/10.1080/ 13547860701252371

Stock, J. H., & Watson, M. W. (2012). *Disentangling the channels of the 2007-2009 recession* (Tech. Rep.). National Bureau of Economic Research. doi: https://doi.org/10.3386/w18094

Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal*, 128 (610), 917–948. doi: https://doi.org/10.1111/ecoj.12593

Trung, N. B. (2019). The spillover effects of us economic policy uncertainty on the global economy: A global var approach. *The North American Journal of Economics and Finance*, 48 , 90–110. doi: https://doi.org/10.1016/j.najef.2019.01.017

Ueda, K. (2010). Determinants of households' inflation expectations in japan and the united states. *Journal of the Japanese and International Economies*, 24 (4), 503–518. doi: https://doi.org/ 10.1016/j.jjie.2010.06.002

Wold, H. (1938). *A study in the analysis of stationary time series* (Doctoral dissertation, Almqvist & Wiksell). Retrieved 12.03.2022, from http://urn.kb.se/resolve?urn=urn:nbn:se:su: diva-72317

WorldBank. (2022). *Global economic prospects*, january 2022. The World Bank. Retrieved 04.05.2022, from http://hdl.handle.net/10986/36519

Zhang, D., Lei, L., Ji, Q., & Kutan, A. M. (2019). Economic policy uncertainty in the us and china and their impact on the global markets. *Economic Modelling*, 79 , 47–56. doi: https:// doi.org/10.1016/j.econmod.2018.09.028

A Appendix A: Lag-order selection

Table 6: Lag-order Selection Criteria

Lag	LL	LR	df	FPE	AIC	HQIC	SBIC
0	-1487.036	.	25	1.928510	14.84613	14.87938	14.92830
1	-26.29050	2834.283	25	1.20e-06	0.560104	0.759606	1.05313
2	68.71133	179.6054	25	6.01e-07	-0.136431	0.229322*	0.767458*
3	105.0709	66.93062	25	5.37e-07*	-0.249462*	0.282542	1.065286
4	129.9566	44.57137*	25	5.39e-07	-0.248325	0.449931	1.477282
5	146.6120	29.00196	25	5.87e-07	-0.165294	0.699213	1.971172
6	159.5095	21.81664	25	6.66e-07	-0.044871	0.985887	2.502454
7	177.9131	30.21483	25	7.16e-07	0.020765	1.217775	2.978949
8	197.5739	31.30075	25	7.63e-07	0.073892	1.437152	3.442934

* optimal lag
Endogenous: MPU ln(GDP) ln(WPI) Rate Reer
Exogenous: _cons
This table presents the Lag-order selection criteria for the VECM model in equation 4.

B Appendix B: Variance decomposition of the VECM model

B.1 Variance decomposition of MPU

Table 7: Results of the Variance Decomposition

Variance Decomposition of MPU:					
Period	fevd (1)	fevd (2)	fevd (3)	fevd(4)	fevd(5)
1	1	0	0	0	0
2	.9714446	.01030575	.00570951	.00027585	.01226424
3	.9482614	.01317462	.01963558	.00276018	.01616817
4	.9313359	.01256629	.02970760	.00523315	.02115706
5	.9132844	.01231742	.03840107	.00866680	.02733028
6	.8950224	.01313629	.04669516	.01183402	.03331210
7	.8770004	.01494577	.05349270	.01459082	.03997031
8	.8601666	.01693521	.05889185	.01696745	.04703886
9	.8451530	.01850130	.06336588	.01918366	.05379616
10	.8317771	.01967423	.06718779	.02126261	.06009831
11	.8194355	.02073294	.07054775	.02321554	.06606823
12	.8077344	.02185160	.07356414	.02502346	.07182643
13	.7965229	.02304356	.07629571	.02669481	.07744302
14	.7857912	.02424011	.07879077	.02825391	.08292401
15	.7755395	.02538398	.08109949	.02972992	.08824711

This table presents the estimates for the variance decomposition obtained with Stata. (1) denotes MPU, (2) denotes ln(GDP), (3) denotes ln(WPI), (4) denotes Rate, (5) denotes (REER)

B.2 Variance decomposition of GDP

Table 8: Results of the Variance Decomposition

Variance Decomposition of Ln(GDP):					
Period	fevd (1)	fevd (2)	fevd (3)	fevd(4)	fevd(5)
1	.01372436	.9862756	.00000000	.00000000	.00000000
2	.01067197	.9565995	.00032155	.00503681	.00737013
3	.01976791	.9333076	.00333704	.00814646	.01544098
4	.01636529	.9254616	.00909092	.01073336	.01834886
5	.01193436	.9218651	.01580596	.01235520	.01803940
6	.02836692	.9193678	.02176590	.01357674	.01692260
7	.02528009	.9180453	.02623580	.01458682	.01585202
8	.02268461	.9170054	.02961178	.01558095	.01511729
9	.02057898	.9155677	.03259918	.01656932	.01468481
10	.01883540	.9137671	.03553233	.01751060	.01435453
11	.01736463	.9119044	.03836033	.01835573	.01401495
12	.01610707	.9101998	.04092894	.01909741	.01366680
13	.01501776	.9087214	.04316751	.01974994	.01334342
14	.01406729	.9074240	.04511062	.02033378	.01306429
15	.01323372	.9062418	.04683413	.02086293	.01282740

This table presents the estimates for the variance decomposition obtained with Stata. (1) denotes MPU, (2) denotes ln(GDP), (3) denotes ln(WPI), (4) denotes Rate, (5) denotes (REER)

B.3 Variance decomposition of WPI

Table 9: Results of the Variance Decomposition

Variance Decomposition of ln(WPI):					
Period	fevd (1)	fevd (2)	fevd (3)	fevd(4)	fevd(5)
1	.00384795	.01622040	.9799316	.00000000	.00000000
2	.00346575	.02769643	.9500511	.00117859	.01760811
3	.00639384	.02194697	.9443248	.00345951	.02387492
4	.00496796	.01498800	.9476243	.00521234	.02720743
5	.00356238	.01069766	.9483092	.00649732	.03093347
6	.00297124	.00819775	.9471939	.00760246	.03403467
7	.00271348	.00659833	.9460277	.00867520	.03598529
8	.00261735	.00554195	.9450324	.00971059	.03709771
9	.00261987	.00478844	.9441570	.01066226	.03777242
10	.00270277	.00419901	.9433313	.01149208	.03827480
11	.00284112	.00372402	.9425167	.01219903	.03871913
12	.00299746	.00334011	.9417408	.01280447	.03911716
13	.00314592	.00302801	.9410411	.01333179	.03945320
14	.00327863	.00277115	.9404283	.01379642	.03972553
15	.00339743	.00255582	.9398922	.01420694	.03994756

This table presents the estimates for the variance decomposition obtained with Stata. (1) denotes MPU, (2) denotes ln(GDP), (3) denotes ln(WPI), (4) denotes Rate, (5) denotes (REER)

B.4 Variance decomposition of Policy Repo Rate

Table 10: Results of the Variance Decomposition

Variance Decomposition of Rate:					
Period	fevd (1)	fevd (2)	fevd (3)	fevd(4)	fevd(5)
1	.00511308	.00002497	.0265613	.9683007	.0000000
2	.00433652	.00162292	.0744053	.9195835	.0000518
3	.00287328	.00177958	.1134753	.8789957	.0028761
4	.00475514	.00124557	.1495186	.8384934	.0059873
5	.00577227	.00083991	.1765637	.8078137	.0090105
6	.00644375	.00067720	.1973706	.7837141	.0117943
7	.00692523	.00062933	.2128547	.7654750	.0141158
8	.00733393	.00060985	.2248197	.7513027	.0159339
9	.00763282	.00059822	.2341237	.7402920	.0173532
10	.00785652	.00059622	.2415573	.7315106	.0184794
11	.00801089	.00060413	.2475782	.7244125	.0193942
12	.00811565	.00061799	.2525301	.7185836	.0201526
13	.00818767	.00063304	.2566434	.7137485	.0207874
14	.00823990	.00064648	.2600989	.7096938	.0213210
15	.00827876	.00065783	.2630314	.7062598	.0217722

This table presents the estimates for the variance decomposition obtained with Stata. (1) denotes MPU, (2) denotes ln(GDP), (3) denotes ln(WPI), (4) denotes Rate, (5) denotes (REER)

B.5 Variance decomposition of Real Exchange Rate

Table 11: Results of the Variance Decomposition

Variance Decomposition of Real Exchange Rate:					
Period	fevd (1)	fevd (2)	fevd (3)	fevd(4)	fevd(5)
1	.0750002	.0047428	.0125702	.0126738	.8950130
2	.0584129	.0021845	.0147428	.0108053	.9138545
3	.0505135	.0017423	.0136694	.0127925	.9212824
4	.0602626	.0013327	.0130433	.0141097	.9112517
5	.0698727	.0010544	.0126658	.0152164	.9011908
6	.0771707	.0008767	.0121951	.0158763	.8938813
7	.0842097	.0007551	.0116931	.0163586	.8869835
8	.0906360	.0006611	.0111997	.0166613	.8808419
9	.0961721	.0005860	.0107298	.0168669	.8756452
10	.1009865	.0005258	.0102976	.0169964	.8711936
11	.1051498	.0004771	.0099085	.0170755	.8673891
12	.1087385	.0004371	.0095592	.0171177	.8641476
13	.1118560	.0004036	.0092450	.0171362	.8613592
14	.1145867	.0003748	.0089623	.0171393	.8589369
15	.1169912	.0003498	.0087078	.0171330	.8568182

This table presents the estimates for the variance decomposition obtained with Stata. (1) denotes MPU, (2) denotes ln(GDP), (3) denotes ln(WPI), (4) denotes Rate, (5) denotes (REER)

C Appendix C: Granger-causality

The tested hypotheses on Granger-causality for three uncertainty indices exhibit notable patterns that confirm the insights into the strong dominance of US uncertainty and monetary policy on the Indian MPU. The theory behind Granger-causality is based on the assumption that a cause cannot come after the effect. Generally, if variable x has an influence on variable y, then the former should help to predict the latter. Therefore, the concept of Granger-causality can be formulated as follows: Let $y_{T\hat{+}h|T}$ represent the h-step ahead forecast of variable *y* at time T. *x* Granger-causes *y* if:

$$\text{MSE}(\hat{y}_{T+h|T}|\text{All information at time T})$$
$$< \text{MSE}(\hat{y}_{T+h|T}|\text{All information at time T, excluding}, x_T, \dots, x_1)$$

where MSE is the Mean Squared Error. Analysing Indian MPU, it can be concluded that MPU does not Granger-cause neither US MPU nor US EPU. However, there is a clear unilateral relationship between US MPU and Indian MPU, as well as US MPU and Indian MPU. This is not surprising, as US monetary policy dominates the Indian MPU and overall uncertainty spillovers originating in the US to emerging countries have been confirmed by the literature. In addition to the above-mentioned influences, there is also a bilateral relationship between US MPU and US EPU.

Table 12: Granger causality of Exogenous Variables

Direction of causality	Wald test statistic	p-value
Dependent variable: MPU		
US MPU $\overset{G}{\nrightarrow}$ MPU	15.634	$2.668 \times 10^{-7***}$
US EPU $\overset{G}{\nrightarrow}$ MPU	25.606	$2.777 \times 10^{-11***}$
Dependent variable: US MPU		
MPU $\overset{G}{\nrightarrow}$ US MPU	2.9749	0.0520
US EPU $\overset{G}{\nrightarrow}$ US MPU	3.8882	0.0211^{*}
Dependent variable: US EPU		
MPU $\overset{G}{\nrightarrow}$ US EPU	1.5742	0.2083
US MPU $\overset{G}{\nrightarrow}$US EPU	5.4423	0.0046^{**}

This table presents the Granger-causality, the Wald statistic and the associated p-values for the uncertainty indices Indian MPU (abbreviated as MPU), US MPU, and US EPU, as introduced in section 3. X $\overset{G}{\nrightarrow}$ Y, indicates the hypothesis that variable X does not Granger-cause variable Y. $*$, $**$, and $***$ denote 5%, 1%, and 0.1% significance level respectively.

D Appendix D: Political Reforms

Table 13: Listings of political and economic developments concerned with MPU from 2002-2021

Year	Developments
2002	Little MPU
2003	Rupee in volatile range, uncertainty in state elections, Iraq invasion threat
2004	Parliament elections, RBI hikes interest rates, inflation rises, high global crude prices
2005	Inflationary pressures, RBI hikes interest rates, weakening of rupee
2006	Highest ever deficit, heavy selling by FIIs, rupee depreciation, inflation, RBI hikes short term rates
2007	Hike in CRR, causing reduction of money in the system, India's competitiveness is impacted by volatile rupee, oil prices rising, impact of the US subprime mortgage crisis, uncertainty about FIIs
2008	Uncertainty regarding US mortgage market, liquidity crunch, inflation fears, oil prices surge, RBI hikes CRR and repo rate, fall of rupee after Lehman failure, recession fear
2009	Recovery under volatile markets, fuel prices hike, deflationary concerns, fear of monetary tightening in China
2010	Inflation fears, fuel prices spur volatility, global liquidity crisis, escalating Eurozone debt crisis, RBI hikes short term rates, rupee appreciates slightly
2011	Inflation fears deepen, volatile rupee, rupee depreciates heavily, European debt crisis, RBI hikes interest rates, US Debt Ceiling debate, S&P downgrades US rating, Moody's downgrades Indian banks, slowdown in China
2012	Rupee made a floating exchange rate, rupee drops to all time low, Eurozone concerns, fear of drought, rising oil prices,
2013	Surging current account deficit, Taper Tantrum, high inflation, Fed rate hike uncertainty, rupee volatility, Fed tapering
2014	Inflation slides, uncertainty about elections,US air strikes in Iraq, Russia-Ukraine conflict, hike in US interest rates, rising oil prices
2015	Rate cut fears, swings in rupee valuation, inflation drops, weak monsoon raises uncertainty about rural economy, Greek crisis, China devalues yuan, Fed hikes interest rates
2016	Rupee on new low against the dollar, inflation targeting implemented, RBI cuts policy rate, RBI governor Raghuram Rajan retires, Brexit referendum, higher inflation, Trump election, demonetisation decision
2017	RBI hikes reverse repo, US strikes in Syria, Indonesia-Pakistan border tensions escalacte, fear of rupee swings, low inflation, high fiscal deficit
2018	Fiscal deficit, rising oil prices, rising inflation, uncertainty regarding Union Budget, weak rupee, currency wars, trade tensions between US and China, RBI governor Urjit Patel resigns
2019	Fear of deflation, farm crisis, Brexit deal, RBI cuts repo rate, rupee rallies, fears of US economic recession, global trade wars, Fed hints at QE
2020	US and Iran tensions, geopolitical tensions in West Asia, uncertainty regarding inflation, YES Bank crisis, Covid-19 breakout
2021	Covid-19 uncertainties, cryptocurrencies crash, Evergrande, RBI keeps rates unchanged

The Impact of the Abolition of the Army on Costa Rica's Educational Development

Abstract

This research paper investigates the impact of the abolition of the army in Costa Rica on long-term educational development. Using a pre-post analysis, I find that the abolition is associated with an average increase of 6.3% in primary school enrollment as a percentage of the total population. Additionally, I find that primary school enrollment in the 20 years following demilitarization was on average 1.9% higher relative to a synthetic Costa Rica without abolishment. I suggest that the abolition of the army had a significant and positive contribution to the country's long-term educational outcomes.

Content

1 Introduction

Following the Civil War of 1948, José Figueres Ferrer, the President of the Founding Junta of the Second Republic of Costa Rica, made the landmark decision of abolishing the national army as part of his efforts to establish a more peaceful and democratic society. The country has since relied on its police force and other non-military means of defense. Since the abolishment of the army, Costa Rica has not experienced a major armed conflict. It has also allowed the government to redirect funds that would have been used for the military towards education, health care, and other social programs. The decision to disband the army and invest in social welfare raised questions about whether this would have any impact on the country's long-term development. This research paper seeks to answer these questions by examining the impacts of the abolition of the army on Costa Rica's educational development. By examining the specific case of Costa Rica, I hope to gain insight into the relationship between demilitarization and education in developing countries.

There is very limited literature on the developmental effects of Costa Rica's demilitarization. The most notable work, made by Abarca and Ramírez (2018), suggests that the abolition of the army yielded long term economic benefits for Costa Rica. By using synthetic control methods, they found that after the abolishment of the army, Costa Rica's per capita GDP grew, on average, more than twice as fast as before the abolishment. My research contributes to Abarca and Ram´ırez's work by expanding the scope of their research to include long-term effects on education, a relevant area that lacks exploration.

To estimate the effect of the abolition of the army on long-run educational development, I first perform a pre-post Analysis of primary school enrollment as a percentage of the total population in Costa Rica before and after the army was abolished. By using data from the Montevideo-Oxford Latin American economic history database (Bértola and Rey, 2018) on primary enrollment rates in Costa Rica, I use simple linear regression to model the relationship between primary enrollment and time. I estimate the magnitude of the change in primary enrollment over time to determine whether there was a significant increase in primary enrollment after the army was abolished.

I then complement this analysis by following the approach of Abadie and Gardeazabal (2003) and Abarca and Ramírez (2018) and using synthetic control methods to estimate the effect of Costa Rica's abolition of the army on primary school enrollment. In the absence of a counterfactual, I compare Costa Rica to a synthetic unit of Latin American countries that resembles key economic and social characteristics of Costa Rica.

My estimates from the pre-post Analysis show that the abolition of the army is associated with an average 6.3% increase in primary school enrollment until 1970. Moreover, the synthetic control shows that primary school enrollment after the abolishment grew, on average, 1.9% more than the synthetic unit, which represents Costa Rica without abolition. Thus, my research suggests that demilitarization had positive consequences on Costa Rica's long-run educational development.

In the next section I will present a brief history of the abolition of the army, followed by an overview of research on the topic. In the third section, I will discuss the methods and techniques I used to address my research question. Then, I will describe the data that I used for my analysis. In the fifth section I will summarize and interpret my experiment results, followed by the last section where I will set forth my conclusions.

2 The Abolition of the Costa Rican Army

2.1 History

The abolition of the army in Costa Rica occurred in 1949, when President Jos´e Figueres Ferrer declared the country's official demilitarization through article 12 of the political constitution. This decision was made as part of a broader effort to create a more peaceful and democratic society in Costa Rica (Muñoz, 2014).

Before the abolition of the army, Costa Rica had a long history of political instability and civil conflict. In the 19th and early 20th centuries, the country was ruled by a series of authoritarian leaders who used the military to maintain their power (Rojas-Fonseca, 2020).

In 1948, however, Costa Rica experienced a major political upheaval. The 1948 elections in Costa Rica were held amid political instability and civil unrest. The main candidates were Rafael Ángel Calderón Guardia, a former president, and Otilio Ulate Blanco, a progressive candidate. The elections were marked by fraud and corruption, and both Calderón and Ulate claimed victory. Calderón ultimately declared the elections void and seized power. This led to a period of political uncertainty and a rise in civil unrest. In April 1948, a group of progressive politicians led by José Figueres Ferrer declared war on the Calderón government, leading to a brief civil war. Calderón was forced to resign, and Figueres became the new president of Costa Rica (Muñoz, 2014; Rojas-Fonseca, 2020).

Figueres was determined to create a more peaceful society in Costa Rica. One of his first actions as president was to abolish the country's armed forces and declare Costa Rica a demilitarized zone. Figueres' decision to abolish the army was met with both praise and criticism. Many Costa Ricans saw the move as a bold and necessary step towards creating a more civil society. Others, however, were concerned that the country would be vulnerable to attack without a standing army. Despite these concerns, Figueres remained committed to his decision and the army was officially dissolved on December 1, 1949 (Rojas Aravena, 2018).

Today, Costa Rica remains one of the few countries in the world without a standing army. Since the abolition of the army, Costa Rica has continued to be a leader in promoting peace and democracy in Central America. The country has not experienced any major conflicts or military coups, and has instead focused on building strong institutions and investing in education and health care. Today, Costa Rica is widely regarded as one of the most stable and prosperous countries in the region.

2.2 Research

There are plenty of studies on the negative impact of wars and military expenditure on economic and social growth and development. For instance, Deger and Sen (1995) present through endogenous growth models, how military expenditure in developing countries has a negative impact on the development of human capital. Collier (2006) contributes to these findings by showing how military expenditure and war slow down development. These are just two of many studies that show the negative relationship between militarism and economic growth (Abadie and Gardeazabal, 2003; Bove, Elia, and Smith, 2014; Dunne and Tian, 2016; Lai and Thyne, 2007).

The question of whether demilitarization contributes to economic and social development is an important one, and the case of Costa Rica provides a valuable example. Despite the widely studied effects of military expenditure on growth and the historical significance of Costa Rica's army abolition in 1949, there is very limited research on the long-term

economic and developmental consequences of this event.

Abarca and Ramírez (2018) provide a unique and unprecedented look at Costa Rica's army abolishment. It is the first study to quantitatively investigate the effect of the abolition on Costa Rica's long-run economic growth with respect to other countries in the region. Abarca and Ram´ırez use long time series data and synthetic control estimates to determine whether the abolition of Costa Rica's army had a positive effect on GDP per capita growth. They find that after demilitarization and until 2010, Costa Rica had an annual GDP per capita growth rate of 2.49%, compared to a pre-demilitarization rate of 0.97% - a 1.52% difference that made Costa Rica go from penultimate in the region to second best. Abarca and Ram´ırez's research provides evidence that for Costa Rica, the abolishment of the military yielded positive long-term effects on its economic activity. They show that the effects are robustly explained by a phenomenon that started in 1951, right after the abolition.

A concern that arises from Abarca and Ramírez study, is that their estimated results might arise from other shocks besides the abolition of the army, most notably, the set of social reforms implemented in Costa Rica in the 1940s, which include - the creation of the Costa Rican Social Security Fund (CCSS), the National University (UCR), and the enactment of a Workers Code (Rosenberg, 1981). To address these concerns, they perform the synthetic control method with 1944 as the treatment year to show that the divergence in real GDP growth between Costa Rica and the synthetic unit does not materialize until 1951.

My research contributes to Abarca and Ramírez's work by expanding into an almost unexplored but highly relevant area - education. As Rojas-Fonseca (2020) presents, the abolishment of the army translated into higher investments in education, however, the actual impact on educational outcomes is poorly documented. My research investigates the quantitative impact of Costa Rica's demilitarization on primary school enrollment, a key indicator of educational outcomes.

3 Experimental Design

3.1 Pre-Post Analysis

Pre-post analysis is a statistical method that is used to evaluate the im pact of a treatment or intervention on a particular outcome variable. This method is based on the principles of ordinary least squares (OLS) regression, which is a statistical technique that is used to model the relationship between a dependent variable and one or more independent variables.

To perform OLS pre-post analysis, data on the dependent and independent variables before and after the intervention must be collected. Then, the data is used to estimate the coefficients for the regression model using OLS regression.

The basic idea behind pre-post analysis is to compare the values of the outcome variable before and after the intervention, and to use regression analysis to determine whether the difference in values is statistically significant. Essentially, outcomes before (t = 0) and after (t = 1) the treatment event are compared:

$$Y_{it} = \alpha + \beta T_{it} + \delta X_{it} + \varepsilon_{it}$$

Where Y_{it} is the outcome variable, α is the intercept (i.e. the expected value of Y_{it} when all independent variables are zero), β is the coefficient (the effect on Y_{it}) of the independent variable T_{it}, which represents the treatment period (0 when $t = 0$ and 1 when $t = 1$), δ is the coefficient of the omitted variables X_{it}, and ε_{it} represents the error term.

To estimate the causal effect of the treatment T_{it} on the outcome variable Y_{it}, expectations are taken:

$$\begin{aligned} E[Y_{i1} - Y_{i0}] &= \beta E[T_{i1} - T_{i0}] + \delta E[X_{i1} - X_{i0}] \\ &= \beta + E[X_{i1} - X_{i0}] \end{aligned}$$

The causal effect β is captured when the omitted variable term $\delta E[X_{i1} - X_{i0}]$ is 0, that is, when the omitted variables X_{it} are constant over time ($X_{i0} = X_{i1}$). Unlike a difference-in-difference (DID) analysis, in a pre-post analysis there are no control units, and hence, the unobserved variables for the treated unit must be constant over time.

In the case of my research, the abolition of the army is unlikely to be the only dependent variable impacting primary school enrollment (the outcome variable), before and after 1949. Hence, the assumption that the unobserved variables for Costa Rica are constant over time is improbable. To add robustness to my experiment, I complement the pre-post analysis

with synthetic control methods.

3.2 Synthetic Control Methods

The synthetic control method was originally proposed by Abadie and Gardeazabal (2003). In classic comparative studies, a unit affected by an event or intervention (also called treated unit), is compared to non-affected units (control units). An issue with such studies is that a treated unit might not always be comparable to control units due to underlying differences between them. If that is the case, a simple comparison of outcomes may not only reflect the impact of the treatment but also of other pre-treatment differences between treated and control units. Synthetic control methods address this issue. The method works on the premise that the treated unit can be more accurately approximated by a combination of similar untreated units than by any single untreated unit.

Synthetic control methods address the shortcomings of a regular DID analysis. A DID analysis requires the existence of a control group that is similar to the treatment group in all respects except for the fact that it does not receive the intervention. This is often difficult to find in practice, especially when the intervention is being implemented on a large scale. To tackle this, the synthetic control method creates a synthetic unit - a weighted combination of the control units. The weights are chosen such that the synthetic unit best resembles key characteristics of the treated unit before the intervention. Then, the estimated post-intervention outcomes for the synthetic unit are used to approximate the outcome that would have been observed in the treatment unit in the absence of treatment, to then obtain more accurate estimates of the impact of the intervention (Abadie, Diamond, and Hainmueller, 2011; Firpo and Possebom, 2018).

As Abadie (2021) reports, a credible application of synthetic control methods has relevant data requirements. First, sufficient data on outcomes and outcome predictors for the treated and untreated units is needed. Second, a synthetic control depends greatly on the availability of pre-intervention information. In order to create a reliable synthetic unit, the synthetic control needs to effectively track the path of the outcome variable for all the units in question before the intervention, thus, increased pre-treatment data leads to enhanced model performance. Finally, extensive post-intervention data is needed to allow for a broader and more comprehensive picture of the effects of the intervention across outcomes of interest.

Here is an overview of the synthetic control method application in Abadie and Gardeazabal (2003). Let J be the number of control units , and $W = (w_1, \dots , w_j)$ a $(J * 1)$ vector of weights which sum to one. The scalar $w_j (j = 1, \dots , J)$ represents the weight of region j in the synthetic unit. The weights W are chosen so that the synthetic unit most closely resembles the actual one before treatment. Let X_1 be a $(K * 1)$ vector of pre treatment values of K predictor variables for the treated unit and X_0 be a $(K * J)$ matrix containing the values of the variables for the J control units. Let V be a diagonal matrix and its diagonal elements' values reflect the relative importance of the different predictor variables. The vector of weights W^* is chosen to minimize $(X_1 - X_0 W)' V (X_1 - X_0 W)$ for w_j greater than or equal to 0 and their sum $(w_1 + \dots + w_j)$ equal to 1. Thus, the vector W^* defines the combination of control units which best resemble the treated unit in predictor variables before the treatment. The vector of interest is W^*, the optimal weights, as they tell the 'similarity' of each control region to the treated unit (how much of it was used to create the synthetic unit). The synthetic unit predictor variables provide an indication of how well the weighted combination of control units reproduce the values of predictor variables for the treated unit before treatment.

Finally, let Y_1 be a $(T * 1)$ vector whose elements are the values of the outcome interest for the treated unit during T time periods. Let Y_0 be a $(T * J)$ matrix which contains the values of the same variables for the control units. The goal is to approximate the path of the outcome of interest that the treatment unit would have experienced in the absence of treatment. This counterfactual outcome path is calculated as the outcome of the synthetic unit $Y_{*1} = Y_0 W^*$. Here, the relevant number is the difference or divergence (if any) between the actual outcome of interest Y_1 and the estimated outcome of interest for the synthetic unit Y_{*1}, namely:

$$Y_1 - Y_{*1}$$

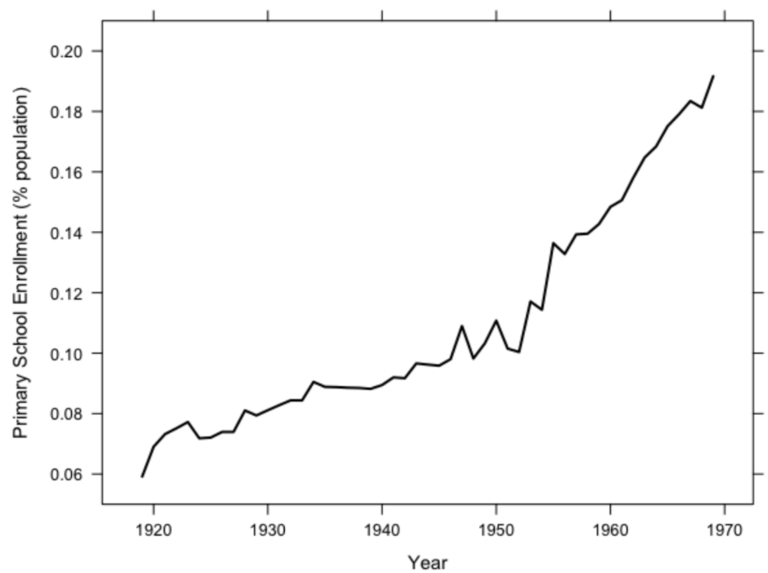
The synthetic control method helps address the issues of lacking an appropriate counterfactual; first, by creating a weighted combination of similar untreated units that better approximate the treated unit (vector W^*), and second, by assigning different weights, and thus different importance, to different predictor variables (matrix V^*). By assigning different weights to different predictor variables, it also tackles the issue of missing data.

4 Data

I used panel data from the Montevideo-Oxford Latin American economic history database (MOxLAD) to perform my analysis. MOxLAD is a collaboration between the Universidad de la República in Montevideo, Uruguay and Oxford University. The database contains a wide range of economic and social data for 20 countries in the region, covering the 20th century up until 2010. It is intended to provide economic and social historians with a comprehensive, single online source of statistical information. The data in MOxLAD has been collected with the aim of providing comprehensive coverage while also ensuring consistency and comparability between countries. The original goal was to cover the 20th century, but it is now being extended both backwards and forwards.

I selected 15 country-specific variables and indicators that I found useful for a pre post analysis and a synthetic control study. I selected the variables based on consistency, comparability and uniformity for every country across the available years, with special emphasis on the country of interest, Costa Rica.

Figure 1: Primary School Enrollment in Costa Rica (1920-1970)



I scraped the data from the online database and created my own panel dataframe made up of 20 units (countries): 1 treated unit (*1 – Costa*

Rica) and 19 control units (*2 – Uruguay, 3 – Argentina, 4 – Bolivia, 5 – Brasil, 6 – Chile, 7 – Colombia, 8 – Cuba, 9 – Dominican Republic, 10 – Ecuador, 11 – Guatemala, 12 – Haiti, 13 – Honduras, 14 – Mexico, 15 – Nicaragua, 16 – Panama, 17 – Paraguay, 18 – Peru, 19 – El Salvador, 20 – Venezuela*). In my sample there are 18 variables (15 country-specific variables, country code, country name, and year) and 100 time periods (1900-2000), for a total of 2020 data entries. Due to lack of data and consistency in the pre-treatment period, I left out Dominican Republic, Haiti, and Paraguay from the experiment sample.

Table 1: Sample Data Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
population	2020	10608.171	21020.9	280	1774	9226.25	174167
land area	2020	999877.8	1871937.156	21040	100521.75	1109372.75	8515767
population density	2020	31.327	48.076	1.268	6.096	32.931	309.117
GDP per capita	1696	2928.883	1959.57	557.413	1543.909	3816.442	11123.585
primary enrollment	1618	0.115	0.044	0	0.077	0.152	0.211
secondary enrollment	1094	0.022	0.025	0	0.004	0.032	0.118
illiteracy	756	26.854	20.989	2.3	10.1	37.9	92
life expectancy	941	60.575	11.461	23.731	55.018	68.961	77.734
unit value of exports	1726	147.736	167.628	10.435	44.259	209.423	1604.275
unit value of imports	1638	161.763	182.232	6.863	51.184	230.79	1929.73
foreign direct investment	720	775.641	2825.904	-889	14	374.75	32779.2
energy consumption per capita	1970	284.453	379.23	0.28	36.505	395.268	2463.36
railway density	1931	0.009	0.01	0	0.002	0.012	0.047
road density	544	0.133	0.137	0.004	0.052	0.146	0.725
vehicles per capita	1215	0.03	0.036	0	0.005	0.035	0.193

Source: MOxLAD database

The selected variable of interest is primary school enrollment as a percentage of the total population, which is a key indicator of educational development. A high value indicates that a country is making progress in providing access to education for all children. Figure 1 illustrates primary school enrollment in Costa Rica from 1920 to 1970.

The selected predictor variables are: population (in thousands), land area (in square kilometers), population density, GDP per capita (in 1990 purchasing power parity USD), secondary school enrollment as a percentage of the total population, illiteracy rates as a percentage of the total population, life expectancy at birth, a fixed weight index of value for both exports and imports (where 1970 = 100), foreign direct investment (in million current USD), energy consumption per capita (in Mtoe), railway and road density, and passenger and commercial vehicles per capita. These variables represent fundamental economic, social, educational, and health characteristics for each of the countries in the panel. Table 1 displays summary statistics for these variables.

As the table shows, most of the variables have high variation, which suggests that the units have very different characteristics and thus, a simple comparison might not capture the true effect of an intervention. This supports the synthetic control method premise that a weighted average of the units can more accurately approximate the treated unit than any single untreated unit. I will expand on this in the next section.

Most of the variables were already available in the MOxLAD database, and for some others I performed the necessary calculations or adjustments. As seen in the table, data availability is not uniform across the units. Furthermore, more economically developed countries tend to have more data availability and reliability. As I explained before, the missing data points should not be a concerning issue due to the design of the synthetic control method, moreover, whenever necessary, I implemented linear interpolation to account for missing data values.

5 Results

5.1 Pre-Post Analysis

5.1.1 Experiment Design

In my pre-post analysis, I evaluate the impact of the 1949 abolition of the army in Costa Rica on primary school enrollment as a percentage of the total population. I perform an OLS regression to model the relationship between the two variables and try to capture, if any, a causal effect.

In my regression equation, $\textit{primary enrollment}_{it}$ is the outcome of interest, and $\textit{treatment}_{it}$ is a dummy variable for the treatment period; it takes on the value 1 for observations collected after the intervention was implemented and the value 0 for ob servations collected before the intervention was applied ($t = 0$ before 1950 and $t = 1$ afterwards). I chose 1950 (the year after the intervention), as the treatment period and narrowed down the available data from 1920 to 1970.

The regression equation is the following:

$$\textit{primary enrollment}_{it} = \alpha + \beta * \textit{treatment}_{it} + \delta X_{it} + \varepsilon_{it}$$

Here, I made the assumption that the omitted variables, represented by X_{it}, are constant over time ($X_{i0} = X_{i1}$), and hence, β, the coefficient of interest, captures the causal effect of the abolition of the army (treatment) on primary school enrollment. I will address the limitations of this approach later in this section.

5.1.2 Findings

Table 2 presents the results of the regression analysis. The dependent variable in the analysis is primary school enrollment, which is measured as a percentage of the total population. The independent variable is a dummy variable for the treatment period, with the value 1 representing the period after the abolition of the army (1950) and the value 0 representing the period before the abolition of the army.

Table 2: The Effect of the Abolition of the Army on Primary School Enrollment

	Dependent variable:
	Primary Enrollment
Treatment	0.063***
	(0.006)
Constant	0.086***
	(0.004)
Observations	51
R^2	0.708
Adjusted R^2	0.702
Residual Std. Error	0.020 (df = 49)
F Statistic	118.793*** (df = 1; 49)
Note:	*p<0.1; **p<0.05; ***p<0.01

Notes: standard errors are in parentheses. The analysis data covers the years 1920 to 1970. The dependent variable (Primary Enrollment) represents primary school enrollment as a percentage of the total population. The independent variable (Treatment) is a dummy variable for the treatment period ($t = 1$ after the abolition of the army (1950), $t = 0$ otherwise).

The table provides estimates of the coefficients for the treatment and constant terms in the regression model, along with measures of the goodness of fit of the model and the statistical significance of the coefficients.
According to the table, the coefficient for the treatment term, β, is 0.063,

which is statistically significant (indicated by the three asterisks after the coefficient estimate). This suggests that, on average, primary school enrollment increased by 0.063 percentage points after the abolition of the army, compared to the period before the abolition of the army. The coefficient for the constant term is 0.086, which is also statistically significant. This represents the expected value of primary school enrollment when the treatment term is equal to 0 (i.e., before the abolition of the army).

Overall, the results of this analysis suggest that the abolition of the army had a statistically significant positive effect on primary school enrollment. The model fits the data well, with an R-squared value of 0.708, indicating that the model explains about 70% of the variation in primary school enrollment. The regression results support my initial hypothesis that the abolition of the army had a positive effect on long-run educational outcomes.

5.1.3 Limitations and Robustness Checks

As I explained before, in this case, the abolition of the army (treatment) is unlikely to be the only dependent variable impacting primary school enrollment (the outcome variable). For instance, other policy reforms and structural changes made both before and after 1950 may also play an important role in influencing primary school enrollment rates. Hence, the assumption that the omitted variables are constant over time ($X_{i0} = X_{i1}$) is unlikely to hold.

As a robustness check, I performed the same regression as before but with 1945 as the treatment year to confirm that the positive estimated effect is less before the intervention of interest was implemented. Table 3 shows the regression results.

Table 3: Robustness Check

	Dependent variable:
	Primary Enrollment
Treatment	0.057***
	(0.007)
Constant	0.083***
	(0.005)
Observations	51
R^2	0.591
Adjusted R^2	0.583
Residual Std. Error	0.024 (df = 49)
F Statistic	70.940*** (df = 1; 49)
Note:	*p<0.1; **p<0.05; ***p<0.01

Notes: standard errors are in parentheses. This table has the same regression design as the one presented in Table 2. The only difference is that the treatment period begins in 1945 (before the army abolition) instead of 1950.

The treatment coefficient is 0.057 for the year 1945, a 0.006 difference compared to the original coefficient of 0.063 in Table 2. This suggests that the 1950 intervention did have a significant and potentially larger effect on primary enrollment than any other treatment before the abolishment of the army.

It is important to note that large-scale policy interventions tend to have lagged effects, and hence, it is difficult to pinpoint the exact moment at which the treatment has an effect on the outcome variable. For this reason, any treatment period selection will prompt questions about arbitrariness.

Despite the limitations of this regression design, which include, the lack of control variables that might influence the dependent variable and the unlikeliness of the omitted variables to stay constant over time, it still

gives us a relevant insight into the effect of the abolition of the army on primary school enrollment. To complement these findings and address some of these limitations, I performed a synthetic control analysis, covered in the next subsection.

5.2 Synthetic Control Method

5.2.1 Experiment Design

Following the approach originally proposed by Abadie and Gardeazabal (2003) and replicated by Abarca and Ramírez (2018), I performed a synthetic control method, as described before, to compare primary school enrollment in Costa Rica to a synthetic unit representing Costa Rica without abolishing its army.

Table 4 compares the pre-abolishment (1920-1950) features of Costa Rica, the synthetic unit, and the sample mean. The synthetic control created this synthetic unit by finding the optimal weights for every unit and variable available in the sample data. Table 5 shows the variable weights (diagonal matrix *V* *), and Table 6 shows the country weights (vector *W* *).

Table 4: Synthetic Control - Pre-Abolishment Charachteristics

	Costa Rica	Synthetic Unit	Sample Mean
population	690.61	1903.20	7088.41
land area	51060.00	187397.68	1216477.50
population density	13.53	16.65	14.70
energy consumption per capita	64.84	138.01	141.32
primary enrollment	0.09	0.09	0.08
gdp per capita	1679.38	2243.05	2013.96
illiteracy	30.73	39.79	51.26
life expectancy	45.73	45.69	39.31
unit value of exports	34.32	43.72	47.27
unit value of imports	50.44	58.44	53.31
foreign direct investment	27.00	31.99	190.31
railway density	0.01	0.01	0.01
road density	0.06	0.11	0.05
vehicles per capita	0.01	0.01	0.01

Notes: this table presents the pre-treatment (1920-1950) characteristics of the treated unit (Costa Rica), the synthetic unit created by the synthetic control, and the sample mean.

As Table 4 shows, most of the pre-treatment variables are very similar between Costa Rica and the synthetic unit, with a few exceptions (population, land area, energy consumption per capita). These predictor variables give an indication of how well the weighted combination of control units reproduce the values of predictor variables for the treated unit before treatment.

Every predictor variable also has an associated weight in the synthetic unit, as shown in Table 5. The weights of each variable represent the relative importance of each in the synthetic unit. The two most important variables are life expectancy, with 21.1%, and population, with 20.6%. These are followed by railway density, illiteracy, and foreign direct investment with 10.8%, 9.9%, and 7.8% respectively. The synthetic control left out road density and unit value of imports from the synthetic unit.

Table 5: Synthetic Control - Variable Weights

Variable	Weights
life expectancy	0.211
population	0.206
railway density	0.108
illiteracy	0.099
foreign direct investment	0.078
primary enrollment	0.068
population density	0.067
unit value of exports	0.052
energy consumption per capita	0.047
land area	0.030
gdp per capita	0.026
vehicles per capita	0.008
road density	0.000
unit value of imports	0.000

Notes: this table represents diagonal matrix $V*$, the weights assigned to each pre-treatment (1920-1950) variable in the synthetic unit. The table is ordered by variable weight.

Table 6 shows the optimal country weights; the 'similarity' of each control region to the treated unit (how much of it was used to create the synthetic unit). Uruguay and Ecuador are the countries with the highest weights, with 39.3% and 31.6% respectively, followed by Panama with 17.1%, and El Salvador with 9.4%. The rest of the countries have zero or almost zero weight in the synthetic unit.

Table 6: Synthetic Control - Country Weights

Country	Weights
Uruguay	0.393
Argentina	0.000
Bolivia	0.000
Brasil	0.000
Chile	0.020
Colombia	0.004
Cuba	0.002
Ecuador	0.316
Guatemala	0.000
Honduras	0.000
Mexico	0.000
Nicaragua	0.000
Panama	0.171
Peru	0.000
El Salvador	0.094
Venezuela	0.000

Notes: this table represents vector $W*$, the weights assigned to each country in the synthetic unit.

Now that the synthetic unit was created, the synthetic method can estimate the path of the outcome variable (primary enrollment) for the synthetic unit, pre and post treatment period. Just like in the pre-post analysis, the treatment year is 1950 and the path is narrowed down to the years 1920-1970. The relevant estimate is the difference or divergence between the path of the treated unit (Y_1) and the synthetic unit ($Y_{*1} = Y_0 W^*$).

5.2.2 Findings

Figure 2 plots primary school enrollment for Costa Rica and the estimates for the synthetic unit. The dotted line is the treatment year. After 1950, there is a clear difference between the outcome variable of interest for Costa Rica and its synthetic unit. The difference in paths suggests that after the abolishment of the army in 1949, there was a considerable positive change on primary school enrollment. Costa Rica experienced a significant increase in primary enrollment while the synthetic unit seems to experience an almost linear and relatively moderate trend. Figure 3 complements Figure 2 by plotting the observed gaps for the outcome variable. Table 7 shows the numerical values of the gaps per year.

Figure 2: Synthetic Control Estimates: Primary School Enrollment Paths (1920-1970)

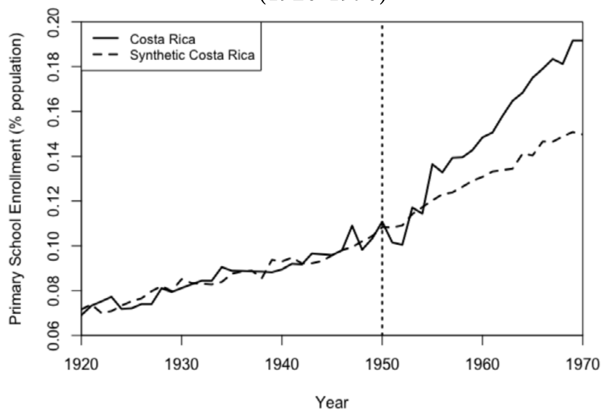


Figure 3: Primary Enrollment Gaps in Costa Rica and Synthetic Unit

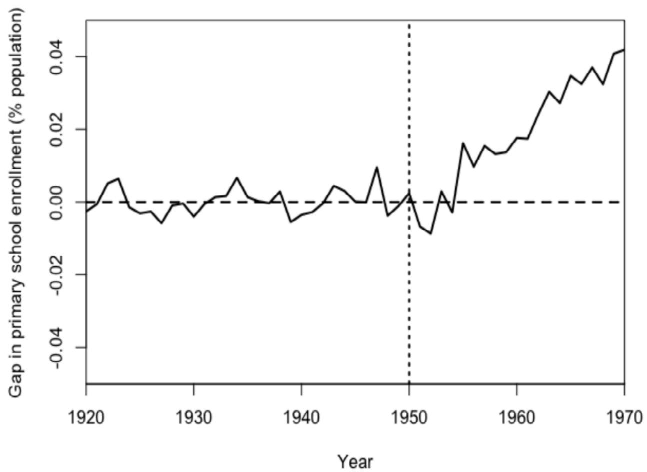


Table 7: Synthetic Control - Primary Enrollment Gaps

Year	Gap	Year	Gap
1920	-0.00	1946	-0.00
1921	-0.00	1947	0.01
1922	0.01	1948	-0.00
1923	0.01	1949	-0.00
1924	-0.00	1950	0.00
1925	-0.00	1951	-0.01
1926	-0.00	1952	-0.01
1927	-0.01	1953	0.00
1928	-0.00	1954	-0.00
1929	-0.00	1955	0.02
1930	-0.00	1956	0.01
1931	-0.00	1957	0.02
1932	0.00	1958	0.01
1933	0.00	1959	0.01
1934	0.01	1960	0.02
1935	0.00	1961	0.02
1936	0.00	1962	0.02
1937	-0.00	1963	0.03
1938	0.00	1964	0.03
1939	-0.01	1965	0.03
1940	-0.00	1966	0.03
1941	-0.00	1967	0.04
1942	-0.00	1968	0.03
1943	0.00	1969	0.04
1944	0.00	1970	0.04
1945	0.00		

Notes: this table shows the gaps in primary school enrollment between Costa Rica and the synthetic unit created by the synthetic control for every year between 1920-1970.

Before the abolition, the average gap is 0.000. This number shows that the synthetic control was successful in creating a synthetic unit that resembles Costa Rica and its primary school enrollment path before treatment. The average gap after treatment is 0.019; in other words, after the abolition, there is an average difference of 1.9% in school enrollment between Costa Rica and its synthetic unit. This suggests that demilitarization is associated with a positive effect on primary school enrollment. As seen in Figure 3 and Table 7, the positive gap in primary enrollment appears after 1955, which might indicate that the effects of the intervention on primary school enrollment were lagged.

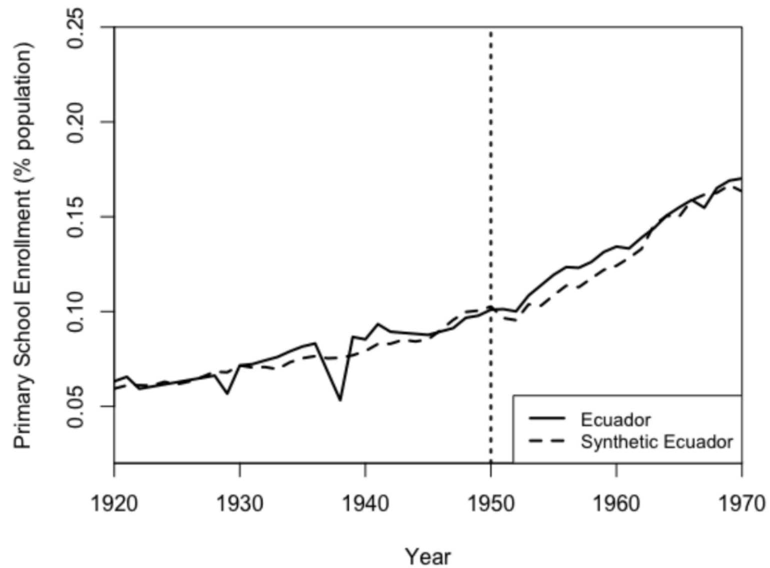
The estimates of the synthetic control provide sufficient evidence to support the hypothesis that, leaving aside other variables, Costa Rica's demilitarization provided long run benefits to educational development. The positive changes in primary enrollment took place not immediately after abolition, but a few years later, which might be expected from a large-scale policy intervention such as this. These results are consistent with the results from the pre-post analysis, and both suggest that demilitarization can have positive effects on long-run educational outcomes.

5.2.3 Limitations and Robustness Checks

There are some relevant limitations in the design of the synthetic control. The first issue is whether the observed gap after treatment is actually due to the abolition of the army or the inability of the synthetic control to reproduce the primary enrollment path in the absence of treatment. To address this, I perform the same method I applied to compute the gap for Costa Rica to another country that did not experience the interven tion. This is called a 'placebo study' (Abadie, Diamond, and Hainmueller, 2011), and its goal is to assess the robustness of the original synthetic control by gauging whether the observed gap corresponds to demilitarization or to some other factor. I chose Ecuador as the placebo unit as it is one of the countries with the highest weights in the synthetic unit.
Figure 4 shows the estimated primary enrollment paths for Ecuador and its synthetic unit. It is clear from the figure that after the treatment period, there is no sizeable gap between the two. In fact, the average gap between Ecuador and synthetic Ecuador in the post-treatment period is 0.004. This suggests that while Ecuador contributed to build Costa Rica's synthetic unit, the observed gap in Costa Rica's case is not driven by a similar event in the control group nor by randomness, but rather by a particular shock that happened in Costa Rica around 1950.

Figure 4: Synthetic Control Placebo: Ecuador

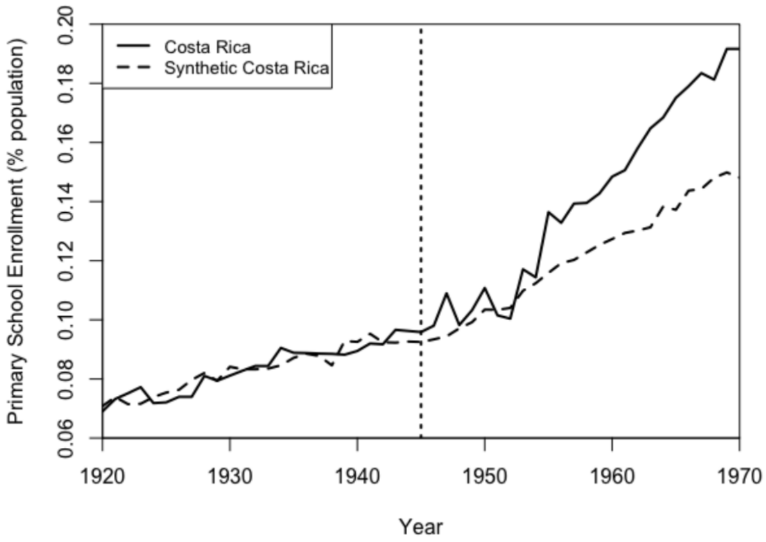
The second limitation of this method is similar to the one addressed in



the pre-post 20 analysis. The gap in primary enrollment might be driven by another intervention or shock in the years when there was no treatment. To control for this, I want to discard that there is a gap between Costa Rica and its synthetic units in years prior to the treatment. Just like I did with the pre-post analysis, I perform the same experiment but with a new treatment year. I chose 1945 as it is before the intervention of interest and because it should capture lagged effects of the 1940s social reforms.

Figure 5 shows the primary enrollment paths for Costa Rica and its synthetic unit with 1945 as the intervention year. Although there is a small positive gap between the two right after 1945, the gap does not become noticeably large until 1955, just like in Figure 2. In addition to this, the average gap from 1945-1949 is 0.005, which is significantly smaller than the mean gap of 0.019 in Figure 2. All of this suggests that there are no positive effects from other shocks on primary enrollment in the pre-treatment period, which provides credibility and robustness to the implemented synthetic control method.

Figure 5: Synthetic Control In-Time Placebo: 1945



Despite its potential limitations, most notably, the observed gap being artificially influenced by unaccounted shocks, and the inability of the synthetic control to exactly reproduce Costa Rica before demilitarization, this method provides a solid approach to establish causal effects and draw inferential conclusions on large-scale policy interventions.

6 Conclusion

Although the relationship between war and economic development has been widely studied, little research has been made on complete demilitarization, especially the case of Costa Rica. This research paper complements the findings of Abarca and Ram´ırez (2018) by studying the effects that the abolition of the army in Costa Rica had on primary school enrollment, a key indicator of educational development, two decades after abolishment (1950-1970). The first experiment of the study, the pre-post analysis, shows that the abolishment is associated with an average increase in primary enrollment of 6.3%. The second experiment, the synthetic control, shows that an average gap of 1.9% appears between the primary school enrollment of Costa Rica and that of an analogous synthetic unit made to represent Costa Rica in the absence of abolishment. Both experiments provide robust evidence to suggest that demilitarization in Costa Rica had positive effects on Costa Rica's long-run

educational development.

The abolition of the army in Costa Rica translated into higher spending in education, health care, and social programs. Funds that would have been directed to military spending, were redirected to construct a more peaceful, democratic, educated, healthier, and civil society. This increased spending on social welfare, made possible by demilitarization, is possibly the driver of the improved educational outcomes reported in this study.

It is important to note that the methodologies covered by this research are far from perfect and have several limitations. First, any improvement in educational outcomes, might be driven by several variables or interventions that were omitted from my experimental design. For instance, the Costa Rican 1940s social reforms might have had lagged effects that would in turn overstate the captured impact of the abolishment. Second, in the case of the synthetic control, the synthetic unit might not be a perfect representation of the treated unit in the absence of treatment. Despite the limitations that the experimental design of this research might have, this study still provides a valuable insight into the relationship between demilitarization and educational outcomes in developing countries.

I hope that this study will serve as a stepping stone in further research on demilitarization in developing countries. The topic of demilitarization is an important and complex issue that requires in-depth analysis. By conducting this study, I hope to contribute to the body of knowledge on Costa Rica's army abolition and provide a foundation for future research on the topic. In particular, I hope that my findings can be used to inform policy decisions on demilitarization and support efforts to promote peace and stability in the region. I believe that demilitarization is crucial for long-term development and prosperity, and I hope that my work will help to further this important goal.

References

Abadie, Alberto. 2021. "Using synthetic controls: Feasibility, data requirements, and methodological aspects." *Journal of Economic Literature 59* (2).

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American statistical Association 105* (490).

_____. 2011. "Synth: An R package for synthetic control methods in comparative case studies." *Journal of Statistical Software 42* (13).

Abadie, Alberto, and Javier Gardeazabal. 2003. "The economic costs of conflict: A case study of the Basque Country." *American economic review 93* (1).

Abarca, Alejandro, and Suráyabi Ramírez. 2018. "A farewell to arms: The Long run developmental effects of Costa Rica's army abolishment." *Observatory of Development at the University of Costa Rica.*

Bértola, Luis, and María Rey. 2018. "The Montevideo-Oxford Latin American economic history database (MOxLAD): origins, contents and sources." *Economic History of Developing Regions 33* (3).

Bove, Vincenzo, Leandro Elia, and Ron P Smith. 2014. "The relationship between panel and synthetic control estimators of the effect of civil war." *Birkbeck Centre for Applied Macroeconomics BCAM Working Papers, Oct.*

Collier, Paul. 2006. "War and military spending in developing countries and their conse quences for development." *The Economics of Peace and Security Journal* 1 (1).

Deger, Saadet, and Somnath Sen. 1995. "Military expenditure and developing countries." *Handbook of defense economics* 1.

Dunne, J Paul, and Nan Tian. 2016. "Military expenditure and economic growth, 1960– 2014." *The Economics of Peace and Security Journal* 11 (2).

Firpo, Sergio, and Vitor Possebom. 2018. "Synthetic control method: Inference, sensitivity analysis and confidence sets." *Journal of Causal Inference* 6 (2).

Lai, Brian, and Clayton Thyne. 2007. "The effect of civil war on education, 1980—97." *Journal of peace research* 44 (3).

Muñoz, Mercedes. 2014. "Costa Rica: La abolición del ejército y la construcción de la paz regional." *Historia y Comunicación Social* 19.

Rojas Aravena, Francisco. 2018. "Costa Rica: siete décadas sin fuerzas armadas." *Nueva Sociedad*, no. 278.

Rojas-Fonseca, Carolina. 2020. "La abolición del ejército, y su entorno Una revisión de las circunstancias y personajes de la crisis del 48."" *Acta Académica* 67 (Noviembre).

Rosenberg, Mark B. 1981. "Social reform in Costa Rica: social security and the presidency of Rafael Angel Calderon." *Hispanic American Historical Review* 61 (2): 278–296.

Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology* 66 (5).

U.S. Public Equity ESG Fund Composite and Parnassus Core Equity Fund: Performance and Factor Attribution

Karthik Nemani, Barrett Buhler

Keywords: ESG, Corporate Financial Performance, Corporate Social Responsibility, Market Efficiency
JEL Classification: G10, G11, G12, G14

Abstract

This is the first paper to examine all U.S. public equity Environmental, Social, and Governance (ESG) funds offered by the Forum for Sustainable and Responsible Investment's (SIF) – for ease of communication, this will be called the ESG Composite – institutional member firms from 2005 to 2020. With a Net Asset Value (NAV) over $150 billion, these funds comprise nearly half of the U.S. public equity ESG investment landscape. The article finds that the ESG Composite maintains performance with the S&P 500 total return index on an overall returns basis with lower volatility, indicating greater risk-adjusted returns. Factor analysis reveals that the ESG Composite returns are primarily driven by underlevered exposure to market returns as well as prevalence of mid-to-large cap and high beta stocks. When isolating the largest fund in the ESG Composite – the Parnassus Core Equity Fund (PRBLX) portfolio – this study finds significant outperformance over the S&P 500 on an overall returns basis. Factor analysis reveals greater emphasis on underleverage to the market and greater preference for large cap, high beta stocks. When compared to the global mutual fund universe, the ESG Composite outperforms in annualized returns and annualized Sharpe ratios, whereas the PRBLX portfolio outperforms in annualized returns, annualized Sharpe ratios, annualized alphas, and annualized information ratios. Conclusions drawn from this study will (1) supplement the discussion on ESG usefulness and (2) present actionable investment insights.

Content

Content

1 Introduction

ESG is a broad term that refers to the consideration of environmental, social, and governance standards into investors' decisions for portfolio selections. Generally, ESG companies seek to generate positive societal byproducts as assessed by nonfinancial data such as carbon emissions, employee satisfaction, and board structure (Armstrong, 2020).

In the past several years ESG investing in the U.S. public equity market increased exponentially, surpassing $380 billion in net assets in 2021 (Lev, 2021).

Such rapid popularity was accompanied by heavy controversy, and ESG bans across the U.S. Florida Governor Ron DeSantis passed anti-ESG legislation in July 2022 that prohibited "State Board of Administration (SBA) fund managers from considering ESG factors when investing the state's money" (Spectrum News Staff, 2022). Also in August, Texas Republican Comptroller Glenn Hegar released a list of 10 companies and 348 investment funds – including BlackRock, Credit Suisse, and UBS – that were barred from doing business with the state. A year prior, Texas enacted legislation prohibiting most state agencies and local government from contracting with such firms (Freedman, 2022).

The primary reason for such bans is the belief that the costs to financial returns outweigh the societal benefits of ESG investing. Florida anti-ESG legislation claims "…the rise of ESG investing [which] sacrifices returns at the altar of…woke agendas," referring to ESG standards as "woke". Furthermore, they state ESG investing "[drives] up costs for consumers in the name of diversity and [sidelines] hardworking Americans by threatening their livelihoods" (Spectrum News Staff, 2022). The Texas Republican Comptroller said, "The environmental, social and corporate governance movement has produced an opaque and perverse system in which some financial companies no longer make decisions in the best interest of their shareholders or their clients," in a statement (Freedman, 2022).

This paper seeks to assess the validity of such statements concerning ESG by comparing the financial returns of the U.S. public equity ESG funds offered by the Forum for Sustainable and Responsible Investment's (SIF) – the ESG Composite – to the S&P 500 total return index on several metrics including cumulative return, annualized return, and Sharpe ratio. It then compares the largest individual fund comprising the ESG

Composite – the PRBLX portfolio – to the S&P 500 total return index on the same metrics. Next, it compares the ESG Composite and the PRBLX portfolio to the global mutual fund universe on annualized returns, annualized Sharpe ratios, annualized alphas, and annualized information ratios. Lastly, it conducts a factor analysis of the ESG Composite and the PRBLX portfolio to draw investment insights.

This paper proceeds as follows: section 3 reviews the literature, section 4 presents the data and empirical strategy, section 5 reviews the results, section 6 concludes.

2 Literature Review

2.1 Establishing the Link Between ESG and Financial Performance

There are numerous papers that study the link between ESG performance and financial performance. For instance, Friede et al. (2015) use evidence from over two-thousand studies of ESG and financial performance and found that 90% of these studies contain a nonnegative relationship that remains approximately the same over time. While this study does find that there is a more positive relationship between ESG and the financial performance of bonds, it does not deny the fact that a positive relationship between ESG and the financial performance of equities exists.

Whelan et al. (2022) build upon Friede et al. (2015) by aggregating over one-thousand studies written between 2015 and 2020. In the corporate studies primarily focused on financial performance, they found that at least 58% of them found a positive relationship between ESG and financial performance. In studies focused on risk-adjusted metrics, 33% of them found a relationship, 26% found a neutral relationship, 28% found mixed results, and only 14% found a negative relationship.

2.2 Social Impact Hypothesis

The social impact hypothesis posits that higher levels of corporate social responsibility (CSR) leads to improved financial performance. This relationship is suggested in the instrumental theories of Garriga and Melé (2004), including the well-known stakeholders' theory, which states that

corporations should strive to do right by all of their stakeholders (including employees, customers, suppliers, local communities, environmental groups, and governmental groups) to achieve true lasting success. Stakeholders' theory is diametrically opposed to shareholders' theory, which states that a company's sole motivation should be to advance its shareholders' interests (McAbee, 2022). The social impact hypothesis believes CSR procures financial performance by creating competitive advantages in the market Jain et al. (2017) improving reputation (Fombrun and Shanley, 1990), building brand image (Murray and Montanari, 1986), and strengthening legitimacy (Hart and Christensen, 2002). Particularly in terms of reputation, Cornell and Shapiro (1987) find that when a company ignores the preferences of interest groups it damages its own reputation, which inversely increases risk premium and overall financial risk. On the other hand, Cornell and Shapiro (1987) maintain that the cost of CSR is almost negligible to its potential benefits.

Most outstanding literature review supports the social impact hypothesis, such as Griffin and Mahon (1997), which found that 33 out of 51 reviewed studies describe a positive correlation between CSR and financial performance. Following this trend, Frooman (1997) found that companies deemed to be irresponsible in their social policies obtained lower profits. Orlitzky et al. (2003) obtained similar results when conducting a meta-analysis of over 50 studies between 1970 and 1997, confirming a positive relationship between socially responsible behavior and financial performance. However, Godfrey et al. (2009) noted that the reason for a positive correlation varied between results, such as positive effect of reputation, or the different methods of measuring CSR and financial performance. Adding to the supportive findings of Orlitzky et al. (2003), Allouche and Laroche (2005) found in an analysis of 82 studies spanning the U.S. and the U.K. that CSR has a positive effect on financial results, with a greater effect measured in the U.K. Tang et al. (2012) also validated the social impact hypothesis, but only when the CSR is adopted as a consistent strategy. In emerging economies, Mishra and Suar (2010) find that CSR strategies prioritizing stakeholders' theory can be profitable to Indian firms. Hebb et al. (2016) revealed empirical evidence about the positive relationship between CSP and aspects such as degree of CSR awareness and stakeholder pressure in Spain. Therefore, it can be concluded that there exists a positive relationship between CSR and financial performance, where CSR is the driving force or independent variable of the relationship.

2.3 Supply and Demand Hypothesis

The supply and demand hypothesis posits that there is no clear link between social and financial performance, as pointed out by McWilliams and Siegel (2001).

Roman et al. (1999) find support for this hypothesis in just 14 of 52 studies reviewed dealing with this relationship. Margolis and Walsh (2003) found evidence for a weak relationship between CSR and financial results in an analysis of 127 studies, in which 31 found it to be either absent or nonsignificant. Van Beurden and Gossling (2008) found nine studies with neutral results, including those by Bowman (1978), Aupperle et al. (1985), Freedman and Jaggi (1986), Fombrum and Shanley (1990), Ruf et al. (2001), and Seifert et al. (2004). There were also studies that found a relationship but reached contradictory conclusions, finding that the relationship is either indeterminate or neutral, according to whether it is positive or negative. Griffin and Mahon (1997) found 9 studies with mixed results out of 51, and in the work of Margolis and Walsh (2003) there were 23 out of 27.

2.4 Trade-Off Hypothesis

According to the trade-off hypothesis, higher CSR levels lead to lower financial performance. Friedman (1970) argues that businesses have no responsibilities other than achieving the highest possible profits, so investing in CSR involves an extra cost that placed a company at a disadvantage in relation to its competitors and also brings in lower profits.

Very few authors found a negative relationship between CSR and financial results in their investigations. Some of the most important empirical studies that did so were those by Brammer et al. (2006) and Van der Laan et al. (2008).

2.5 Available Resources Hypothesis

The available resources hypothesis links good financial performance with high levels of CSR. According to Waddock and Graves (1997), good financial results mean that money can be invested in CSR, so that high profits could be a good indicator of subsequent good social results.

Of the above-mentioned reviews, that by Margolis and Walsh (2003) is the one that concentrates most on studies that consider social responsibility as a dependent variable. Of a total of 22 of this type, 16 found a positive correlation, i.e., good financial performance leads to the adoption of CSR; 3 found the correlation to be nonsignificant, and a further 3 found it to be bidirectional. Studies such as those by McGuire et al. (1988, 1990) provide empirical support for this hypothesis.

2.6 Available Resources Hypothesis

The managerial opportunism hypothesis considers that higher financial performance levels lead to lower CSR levels. Authors such as Person and O'bannon (1997) argue that directors may act to increase their personal benefits and reduce investment in CSR when profit levels are high. Similarly, if profits are low, directors may attempt to justify the situation by blaming ambitious social programs.

This hypothesis is empirically validated in the work of Posner and Schmidt (1992).

2.6 Gaps in Literature

Although a substantial number of studies show a positive relationship between financial performance and individual companies exhibiting CSR strategies, few look at the landscape of ESG funds, which compile such companies to build an entire portfolio. Furthermore, the scope of "positive financial performance" is loosely defined and often differing in many studies, with no standard benchmark for returns to be compared with. Some studies have compared ESG funds with a benchmark, but these funds also hold international equities or bonds and inaccurately compare them to the U.S. public equity-based S&P 500. This study compares a list of 70 U.S. public equity ESG funds to the S&P 500 to maintain the "apples-to-apples" theme and generate tangible, consistent metrics of performance.

3 Data and Empirical Strategy

3.1 U.S. SIF

The ESG Composite is formed by filtering public equity ESG funds of fered by the U.S. SIF member firms. The U.S. SIF is the leading voice in advancing sustainable investing across all asset classes with the mission to "rapidly shift investment practices toward sustainability, focusing on long-term investment and the generation of positive social and environmental impacts." Institutional members of the U.S. SIF manager $5 trillion in assets under management (AUM), and include investment management firms, advisory firms, mutual fund companies, assets owners and broker-dealers, among others. The U.S. SIF is supported by the U.S. SIF Foundation, a 501(c)(3) nonprofit organization that seeks to educate, research and propel the mission of U.S. SIF (US SIF, 2022)

3.2 ESG Composite

The ESG Composite was created by filtering U.S. SIF Sustainable Investment Mutual Funds and ETFs Chart to all U.S. public equity ESG funds on the database.

3.3 ESG Composite, PRBLX, & Mutual Funds Returns Data

Annual total returns of over 23,000 active equity mutual funds through the year 2020 were scraped from YahooFinance and accumulated into one dataset, linked here: https://www.kaggle.com/datasets/stefanoleone992/m utual-funds-and-etfs/versions/3?resource=download.

This dataset is available to the public, along with https://finance.yahoo.com. This dataset was filtered for over 10,000 active equity mutual funds that were operative during some period between January 2005 and December 2020. The ESG Composite was further filtered from this list to the 70 funds from on the U.S. SIF database, and annual total returns of each fund from 2005 to 2020 were provided.

Annual total returns data for the PRBLX portfolio were provided by Yahoo finance, linked here: https://finance.yahoo.com/quote/PRBLX/performanc e?p=PRBLX.

Annual total returns data for the S&P 500 total return index were provided by YahooFinance, linked here: https://finance.yahoo.com/quote/%5EGSPC/history? p=%5EGSPC.

Methods for calculating returns can be found in the appendix.

3.4 Calculating ESG Composite Returns

Unlike the PRBLX portfolio and S&P 500, the ESG Composite was a list of funds. A simple average or median of list returns were susceptible to high volatility from small funds, so a weighted average based on NAV was used. The total NAV of the ESG Composite was calculated by summing each NAV, and then a proportion was calculated by dividing fund-specific NAV by the sum. Finally, the proportion was multiplied by annual total return for each fund per year, and year-specific values were summed to create NAV-weighted annual total returns for the ESG Composite.

3.5 Factor Data and Analysis

In factor analysis of the ESG Composite and PRBLX portfolio excess returns (Rp,t - Rf,t), several specifications and regressions are provided using popular academic factors, with data from January 2005 to December 2020. This includes the CAPM (Equation 1) regressing the ESG Composite and PRBLX portfolio excess returns on a leverage factor (MKT-Rf) defined by the S&P 500 minus the risk-free 3-month T-bill rate:

$$(R_{p,t} - R_{f,t}) = \alpha + \beta \times (R_{m,t} - R_{f,t}) + \varepsilon_t \qquad (1)$$

Analysis also includes a Fama-French (1993) Three Factor Model (Equation 2) that regresses the ESG Composite and PRBLX portfolio excess returns on a leverage factor (MKT-Rf) in addition to size (SMB) and value (HML) factors obtained from the Ken French data library:

$$(R_{p,t} - R_{f,t}) = \alpha + \beta_{MKT} \times (R_{m,t} - R_{f,t}) + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t + \varepsilon_t \quad (2)$$

Another specification is provided using the Four Factor Model (Equation 3) that includes a momentum factor (UMD), also obtained from the Ken French data library:

$$(R_{p,t} - R_{f,t}) = \alpha + \beta_{MKT} \times (R_{m,t} - R_{f,t}) + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t + \beta_{UMD} \times UMD_t + \varepsilon_t \qquad (3)$$

In separate specifications, this study also regresses the PRBLX portfolio excess returns on the Frazzini and Pedersen (2014) Betting-Against-Beta factor and the Asness et al. (2013) Quality Minus Junk (QMJ) factor.

3.6 Synthetic Portfolio Construction

Systematic synthetic portfolios are constructed from the same regressions of monthly returns in Table 3 and Table 4, namely the Four Factor regression using data over the entire time period (January 2005 to December 2020). The portfolio is rebalanced annually at year-end to keep constant weights. The explanatory variables are the monthly returns of the standard size, value, and momentum factors.

4 Results

4.1 Versus S&P 500

Table 1 below displays the side-by-side performance of the ESG Composite and the S&P 500 total return index.

Table 1. Summary Statistics for the ESG Composite Versus the S&P 500

Time Period	ESG Composite Return	S&P 500 Total Return
2005	2.47%	4.83%
2006	10.03%	15.61%
2007	5.58%	5.49%
2008	-26.20%	-36.55%
2009	26.71%	25.93%
2010	13.83%	14.82%
2011	-0.31%	2.10%
2012	13.78%	15.89%
2013	30.26%	32.04%
2014	11.38%	13.52%
2015	-0.53%	1.37%
2016	10.87%	11.76%
2017	19.57%	21.60%
2018	-3.63%	-4.23%
2019	30.77%	31.19%
2020	21.18%	18.05%

Summary Statistics	ESG Composite Return	S&P 500 Total Return
Cumulative Return (01/01/2005-12/31/2020)	319.19%	325.06%
Annualized Return	10.04%	10.13%
Standard Deviation	14.51%	16.40%
Downside Deviation	12.45%	22.85%
Sharpe Ratio	0.60	0.54
Sortino Ratio	0.70	0.39
Active Return (vs. S&P 500 Total Return)	-0.09%	
Tracking Error (vs. S&P 500 Total Return)	3.45%	
Information Ratio (vs. S&P 500 Total Return)	-0.03	

On cumulative return, the ESG Composite is less than the S&P 500, at 319.91% and 325.06%, respectively. Annualized return of the ESG Composite is approximately equal to the S&P 500, with a difference of 0.09%. Standard deviation of the ESG Composite is less than the S&P 500, with a difference of 1.89%. Downside deviation of the ESG Composite is much lower than the S&P 500, with a difference of 10.4%. Sharpe ratio of the ESG Composite is slightly greater than the S&P 500, with a difference of 0.06. However, Sortino ratio of the ESG Composite is nearly double the S&P 500, with a difference of 0.31. Active return and Information Ratio of the ESG Composite are both negative, but tracking error is relatively low at 3.45%.

Figure 1 below displays the ESG Composite and S&P 500 cumulative returns tracked from 2005 to 2020.

Figure 1. ESG Composite Versus S&P 500 Cumulative Returns

Table 2 below displays the side-by-side performance of the PRBLX portfolio and the S&P 500 total return index.

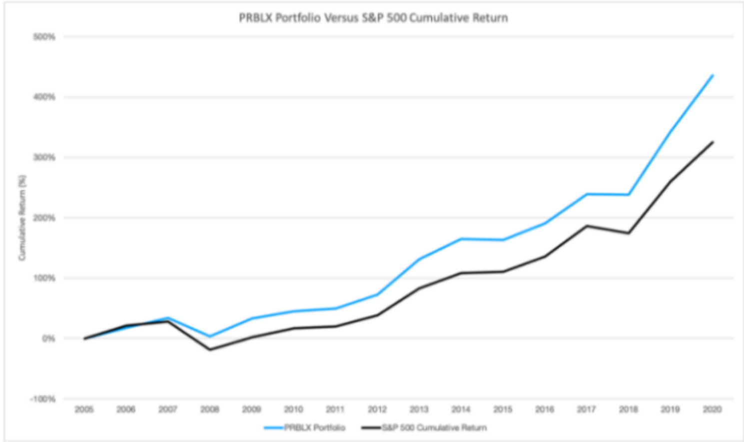
Table 2. Summary Statistics for the PRBLX Portfolio Versus the S&P 500

Time Period	PRBLX Portfolio Return	S&P 500 Total Return
2005	2.62%	4.83%
2006	14.70%	15.61%
2007	14.13%	5.49%
2008	-22.96%	-36.55%
2009	28.75%	25.93%
2010	8.87%	14.82%
2011	3.13%	2.10%
2012	15.43%	15.89%
2013	33.98%	32.04%
2014	14.49%	13.52%
2015	-0.55%	1.37%
2016	10.41%	11.76%
2017	16.58%	21.60%
2018	-0.18%	-4.23%
2019	30.69%	31.19%
2020	21.19%	18.05%
Summary Statistics	**PRBLX Portfolio Return**	**S&P 500 Total Return**
Cumulative Return (01/01/2005-12/31/2020)	436.10%	325.06%
Annualized Return	11.84%	10.13%
Standard Deviation	14.04%	16.40%
Downside Deviation	10.65%	22.85%
Sharpe Ratio	0.75	0.54
Sortino Ratio	0.99	0.39
Active Return (vs. S&P 500 Total Return)	1.72%	
Tracking Error (vs. S&P 500 Total Return)	4.84%	
Information Ratio (vs. S&P 500 Total Return)	0.35	

The PRBLX portfolio has a significantly greater cumulative return than the S&P 500, at 436.10% versus 325.06%, respectively. Annualized return of the PRBLX portfolio is also greater than the S&P 500, with a difference of 1.71%. Standard deviation of the PRBLX portfolio is less than the S&P 500, with a difference of 2.36%. Downside deviation of the ESG Composite is significantly lower than the S&P 500, with a difference of 12.2%. Sharpe ratio of the ESG Composite is greater than the S&P 500, with a difference of 0.21. Furthermore, Sortino ratio of the PRBLX is over double the S&P 500, with a difference of 0.6. Active return and Information Ratio of the PRBLX portfolio are both positive, indicating outperformance over the S&P 500. Tracking error is still low at 4.84%.

Figure 2 below displays the PRBLX portfolio and S&P 500 cumulative returns tracked from 2005 to 2020.

Figure 2. PRBLX Portfolio Versus S&P 500 Cumulative Returns



4.2 Versus Global Mutual Fund Universe

Figure 3 below shows where the ESG Composite (vertical red dashed line) compares against the distribution of annualized returns, Sharpe ratios, alphas, and information ratios of all actively managed equity funds operative between 2005 and 2020.

Figure 3. ESG Composite Compared to YahooFinance Mutual Fund Universe (Jan. 1, 2005 – Dec. 31, 2020)

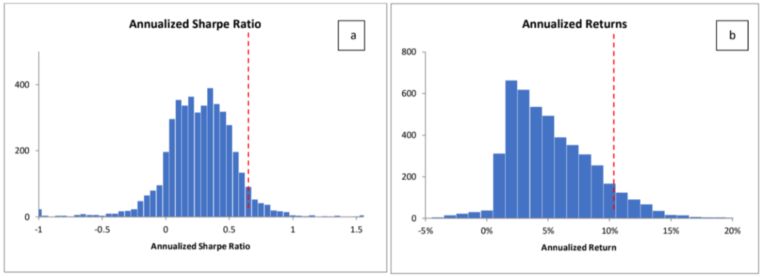


Figure 3a shows the ESG Composite in the higher ranges of annualized Sharpe ratio amongst the mutual fund universe. Figure 3b shows the ESG Composite in the higher ranges of annualized returns amongst the mutual fund universe.

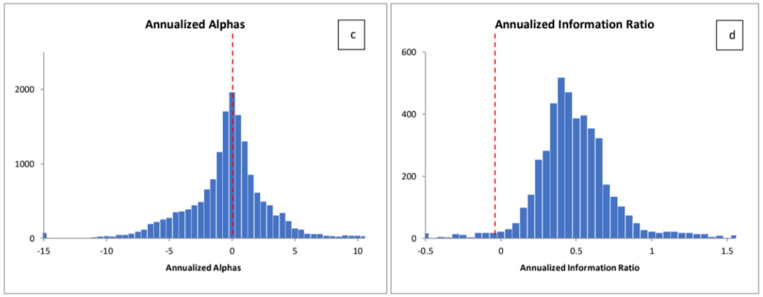


Figure 3c shows the ESG Composite near the center of the annualized alphas amongst the mutual fund universe. Figure 3d shows the ESG Composite near the lower ranges of annualized information ratios amongst the mutual fund universe.

Figure 4 below shows where the PRBLX portfolio (vertical black dashed line) compares against the distribution of annualized returns, Sharpe ratios, alphas, and information ratios of all actively managed equity funds operative between 2005 and 2020.

Figure 4. PRBLX Portfolio Compared to YahooFinance Mutual Fund Universe (Jan. 1, 2005 – Dec. 31, 2020)

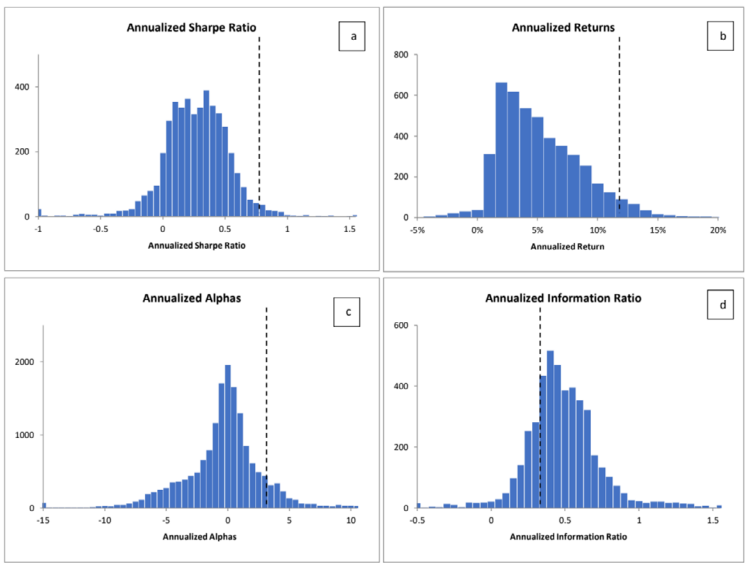


Figure 4a shows the PRBLX portfolio in the higher ranges of annualized

Sharpe ratio amongst the mutual fund universe. Figure 4b shows the PRBLX portfolio in the higher ranges of annualized returns amongst the mutual fund universe. Figure 4c shows the PRBLX portfolio near the center of the annualized alphas amongst the mutual fund universe. Figure 4d shows the PRBLX portfolio near the lower ranges of annualized information ratios amongst the mutual fund universe.

4.3 Factor Attribution

Table 3 below displays the results of factor regression of the ESG Composite.

Table 3. ESG Composite Exposures: What Kind of Companies Do U.S. Public Equity ESG Funds Own?

	CAPM	Fama-French (1993)	Carhart (1997)	Frazzini-Pedersen (2014)	Asness-Frazzini-Pedersen (2013)
Alpha	0.497% (0.491)	0.074% (0.932)	0.283% (0.754)	0.226% (0.716)	-0.283% (0.738)
MKT-R_f	**0.834***** (~0)	**0.868***** (~0)	**0.850***** (~0)	**0.951***** (~0)	**0.979***** (~0)
SMB		-0.066 (0.543)	-0.097 (0.397)	**-0.163*** (0.063)	-0.123 (0.201)
HML		0.031 (0.376)	0.026 (0.457)	0.013 (0.601)	0.002 (0.955)
UMD			-0.031 (0.338)	-0.027 (0.235)	-0.030 (0.196)
BAB				**-2.218***** (0.004)	**-2.242***** (0.005)
QMJ					0.930 (0.381)

Alpha values for the Capital Asset Pricing Model (CAPM), Fama-French, Carhart, Frazzini-Pedersen, and Asness-Frazzini-Pedersen equations are statistically insignificant. Thus, no conclusions can be drawn from them. Traditional leverage factor betas (MKT- Rf) are statistically significant on the 1% scale across all equations, ranging from 0.834 to 0.979. Small minus big (SMB) factor beta is statistically significant at the 10% scale, with a negative value of -0.163. Betting-against-beta (BAB) factor betas are statistically significant at the 1% scale in both equations that incorporate them, with values less than -2.2.

Table 4 below displays the results of factor regression of the ESG Composite.

Table 4. PRBLX Portfolio Exposures: What Kind of Companies Does the PRBLX Portfolio Own?

	CAPM	Fama-French (1993)	Carhart (1997)	Frazzini-Pedersen (2014)	Asness-Frazzini-Pedersen (2013)
Alpha	**2.66%**** (0.039)	2.26% (0.133)	**2.66%*** (0.086)	**2.60%*** (0.061)	1.64% (0.357)
MKT-R_f	**0.779***** (~0)	**0.815***** (~0)	**0.778***** (~0)	**0.898***** (~0)	**0.950***** (~0)
SMB		-0.195 (0.284)	-0.255 (0.177)	**-0.332*** (0.063)	-0.258 (0.194)
HML		0.018 (0.744)	0.009 (0.871)	-0.007 (0.8916)	-0.027 (0.625)
UMD			-0.061 (0.247)	-0.056 (0.2268)	-0.063 (0.196)
BAB				**-2.609*** (0.0617)	**-2.655*** (0.064)
QMJ					1.748 (0.421)

Alpha values for the CAPM, Carhart, and Frazzini-Pedersen equations are statistically significant, with values around 2.6%. Traditional leverage factor betas (MKT- Rf) are statistically significant on the 1% scale across all equations, ranging from 0.779 to 0.950. Small minus big (SMB) factor beta is statistically significant at the 10% scale, with a negative value of -0.332. Betting-against-beta (BAB) factor betas are statistically significant at the 1% scale in both equations that incorporate them, with values less than -2.6.

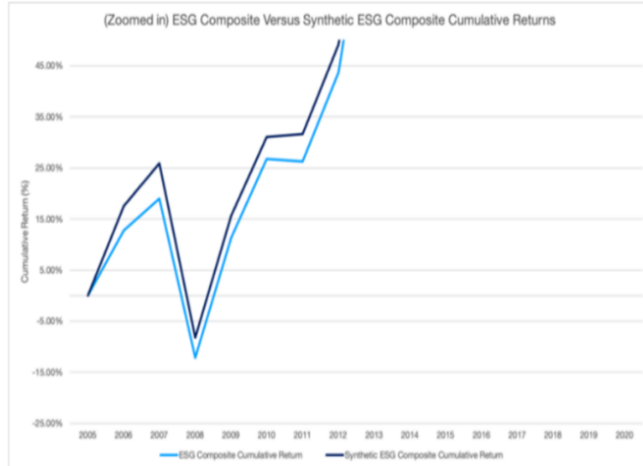
4.4 Synthetic Portfolio Adjustment

Figure 5 below shows calendar-time returns of a synthetic portfolio of the ESG Composite that uses the factor loadings as estimated from factor regression analysis.

Figure 5. ESG Composite vs. Synthetic ESG Composite Cumulative Returns

Figure 6 below shows a zoomed view of the calendar-time returns of the synthetic ESG Composite portfolio.

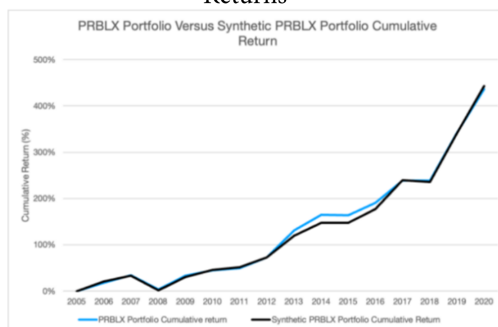
Figure 6. (Zoomed in) ESG Composite vs. Synthetic ESG Composite Cumulative Returns



The synthetic ESG Composite slightly outperforms the ESG Composite's actual cumulative returns for the entire period, particularly between 2005 and 2012, with an approximately 6% improvement in 2007 and an approximately 5% improvement in 2010.

Figure 7 below shows calendar-time returns of a synthetic portfolio of the PRBLX portfolio that uses the factor loadings as estimated from factor regression analysis.

Figure 7. PRBLX Portfolio vs. Synthetic PRBLX Portfolio Cumulative Returns



The synthetic ESG Composite slightly underperforms the PRBLX portfolios actual cumulative returns for the entire period, particularly between 2005 and 2015, with an approximately 10% decrease in 2013 and an approximately 15% decrease in 2015.

5 Interpretation

5.1 ESG Composite

Due to statistically insignificant differences in cumulative return and annualized return, it can be assumed that the ESG Composite produces approximately equal returns as the S&P 500. This is further proven in figure 1, as there is little deviation between cumulative returns at any point in the period of analysis. In a worst-case scenario, the ESG Composite minimally underperforms, as active return and information ratio are only slightly negative. Ultimately, it can be concluded that both U.S. public equity ESG funds and the S&P 500 will produce a return of approximately 10% per year. On the other hand, volatility of the ESG Composite is significantly less than the S&P 500, as shown by the lower standard deviation and downside deviation. The S&P 500's particularly high downside deviation implies that, compared to U.S. public equity ESG funds, investing in the S&P 500 produces much greater risk of negative returns. Lower volatility combined with equal returns means the ESG Composite produces greater risk-adjusted returns. This is supported by the fact that Sharpe ratio and Sortino ratio of the ESG Composite are higher than the S&P 500. The ESG Composite's particularly high Sortino ratio implies that, compared to the S&P 500, investing in U.S. public equity ESG funds produces much greater risk adjusted returns.

Figure 3 shows the ESG Composite in the higher ranges of annualized Sharpe ratio and annualized returns of the global mutual fund universe. This implies that U.S. public equity ESG funds generally outperform mutual funds from both a returns and risk-adjusted returns basis. Annualized alphas of the ESG composite are average compared to the mutual fund universe, indicating that U.S. public equity ESG funds offer just as much outperformance over the market as the average mutual fund. Supplementing average to above-average returns with exceptional risk-adjusted returns implies that U.S. public equity ESG funds offer much more stability than their mutual fund competitors.

As described in the introduction and literature review, the primary rea-

son for ESG bans in the U.S. is the concern of less returns. This analysis dissuades such sentiment by proving that U.S. ESG funds produce greater risk-adjusted returns than the market itself.

Factor analysis of the ESG Composite find statistically insignificant alphas across all equations, suggesting that U.S. public equity ESG funds do not deviate much in returns from the S&P 500. Across all specifications, the results demonstrate slight leverage (investing in the market portfolio), with traditional leverage factor betas near 1, especially in the Asness Frazzini-Pedersen equation. This is supported by the relatively low tracking error of the ESG Composite, along with figure 1. In the Frazzini-Pedersen equation, the study finds that the ESG Composite has more exposure to large caps given the negative SMB factor. When including the Frazzini and Pedersen (2014) Betting-Against-Beta factor and the Asness et al. (2013) Quality Minus Junk (QMJ) factor, this study finds further evidence that the ESG Composite tilts toward large cap stocks and stocks with high beta exposure.

5.2 PRBLX Portfolio

With significant differences in cumulative return and annualized return, it can be assumed that the PRBLX portfolio produces much greater returns than the S&P 500. This is further proven in figure 2, as the PRBLX portfolio begins deviating from the S&P 500 in 2007
and continues through 2020. Both active return and information ratio are positive as well, indicating that the PRBLX portfolio minimally outperforms in a worst-case scenario. Ultimately, it can be concluded that the PRBLX portfolio will produce greater returns than the S&P 500's return of approximately 10% per year. Still, volatility of the PRBLX portfolio is significantly less than the S&P 500, as shown by the lower standard and downside deviations. The drastic difference between downside deviations of the PRBLX portfolio and the S&P 500 implies that, compared to the Parnassus Core Equity Fund, investing in the S&P 500 produces much greater risk of negative returns. Lower volatility combined with greater returns means the PRBLX portfolio produces exceptionally greater risk-adjusted returns. This is supported by the fact that Sharpe ratio and Sortino ratio are much larger than the S&P 500. The PRBLX portfolio's Sortino ratio is over double the S&P 500, implying that investing in the Parnassus Core Equity fund produces significantly greater risk-adjusted returns than the S&P 500.

Figure 4 shows the PRBLX portfolio in the highest ranges of annualized Sharpe ratio, annualized returns, and annualized alphas of the global mutual fund universe. This implies that U.S. public equity significantly outperform mutual funds from a returns and risk-adjusted basis, as well as exceed mutual funds in their own outperformance over the market. Information ratio of the PRBLX portfolio is near average but still positive, indicating the PRBLX portfolio provides at least as much outperformance as the S&P 500.

All of these results dissuade anti-ESG sentiment in the U.S. spurred by concern over returns by proving that the PRBLX ESG fund produces significantly greater returns and risk-adjusted returns than the global mutual fund universe. Furthermore, it offers specific investment insights that adopting holdings strategies of the PRBLX portfolio can offer the greatest financial benefits of ESG investing.

Factor analysis of the PRBLX portfolio find statistically significant alphas in the CAPM, Carhart, and Frazzini-Pedersen equations of around 2.66%. This implies that the PRBLX portfolio produces expected outperformance over the S&P 500 of at least 2%. Compared to the ESG Composite, traditional leverage factor betas for the PRBLX portfolio are lower, indicating the Parnassus Core Equity fund is less levered to the market than most U.S. public equity ESG funds. On the other hand, SMB value in the same specification is nearly twice as negative the ESG Composite, indicating the Parnassus Core Equity Fund has a much greater preference for large caps than most U.S. public equity ESG funds. Furthermore, BAB factor betas are more negative than the ESG Composite, indicating the Parnassus Core Equity fund has a greater preference for high beta stocks than most U.S. public equity ESG funds.

5.3 Investment Insights

The composite of U.S. public equity ESG funds produced greater risk-adjusted returns than the S&P 500, with slight underleverage to the market, yet a preference for large cap, high beta stocks. However, there is potential for optimization if these preferences are strengthened. Such improvement is shown in the synthetic portfolio construction of the ESG Composite (figure 6), which used factor loadings from ESG Composite regression to create 5-6% greater returns in certain years. A real-life example of this optimization is such through the Parnassus Core Equity Fund, which maximized such outperformance over the S&P 500 with a

more pronounced underleverage to the market, and a stronger preference for large cap and high beta stocks. Synthetic portfolio construction of the Parnassus Core Equity Fund shows that it cannot be optimized any further, as the factor-derived model produced 10-15% worse returns than the actual portfolio in certain years.

As a result, it can be concluded that the most valuable returns in ESG investing come from prioritizing established, high cash flow companies that outperform during periods of economic growth and are stable during contractions. Holdings data of the Parnassus Core Equity Fund support such insight, with companies like Microsoft, Apple, and Alphabet of the largest selections.

6 Conclusion

The models and procedures from the study of 70 U.S. public equity ESG funds and the isolated Parnassus Core Equity ESG Fund provide conclusive empirical evidence that U.S.-based public equity ESG funds produce greater risk-adjusted returns than the market. This counters anti-ESG sentiment claiming U.S. ESG funds produce worse financial returns and builds upon former analysis finding a positive correlation between CSR and financial performance. Factor analysis reveals that preference for large-cap, high beta stocks that outperform during periods of economic expansion will produce the greatest financial returns in the U.S. public equity ESG space, as shown by analysis of the Parnassus Core Equity Fund. This means investing into blue chip, high cash flow companies like Microsoft, Apple, and Alphabet will produce the greatest financial returns while balancing ESG criteria.

The shortcomings of this study extend to data collection procedures and testing methodology. Although the U.S. SIF provides a significant portion of U.S. public equity ESG funds, an analysis of all funds in the space would provide a more accurate representation of the relationship between US public equity ESG funds and the S&P 500. However, creating such a dataset would require significant effort to analyze individual firms' investment processes. Furthermore, most of the tests in factor regression produced statistically insignificant results, which were ignored. Advanced regression analyses would reduce this insignificance, allowing more insights to be drawn.

7 References

Allouche, J., & Laroche, P. (2005). A meta-analytical investigation of the relationship between corporate social and financial performance. *Revue de gestion des ressources humaines*, (57), 18.

Armstrong, A. (2020). Ethics and ESG. *Australasian Accounting, Business and Finance Journal*, 14(3), 6- 17.

Asness, C., & Frazzini, A. (2013). The devil in HML's details. *The Journal of Portfolio Management*, 39(4), 49-68.

Asness, C., Frazzini, A. (2022). The devil in HML's details: factors, monthly [Data set]. AQR Capital Management. https://www.aqr.com/Insights/Datasets/ The-Devil-in-HMLs-Details-Factors-Monthly

Asness, C., Frazzini, A., Pedersen, L. (2022). Quality minus junk: factors, monthly [Data set]. AQR Capital Management. https://www.aqr.com/Insights/Datasets/ Quality-Minus-Junk-Factors-Monthly

Aupperle, K. E., Carroll, A. B., & Hatfield, J. D. (1985). An empirical examination of the relationship between corporate social responsibility and profitability. *Academy of management Journal*, 28(2), 446-463.

Ben Brik, A., Rettab, B., & Mellahi, K. (2011). Market orientation, corporate social responsibility, and business performance. *Journal of Business Ethics*, 99(3), 307-324.

Bowman, E. H. (1978). Strategy, annual reports, and alchemy. *California management review*, 20(3), 64- 71.

Brammer, S., Brooks, C., & Pavelin, S. (2006). Corporate social performance and stock returns: UK evidence from disaggregate measures. *Financial management*, 35(3), 97-116.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1), 57-82.

Cornell, B., & Shapiro, A. C. (1987). Corporate stakeholders and corporate finance. *Financial management*, 5-14.

Download TMUBMUSD03M Data: U.S. 3 month Treasury Bill Price data. MarketWatch. (n.d.). Retrieved January 9, 2023, from https://www.marketwatch.com/investing/bond/tmub musd03m/download-data?countrycode=bx&mod=mw_quote_tab

Garriga, E., & Melé, D. (2004). Corporate social responsibility theories: Mapping the territory. *Journal of business ethics*, 53(1), 51-71.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1), 3-56.

Fombrun, C., & Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. *Academy of management Journal*, 33(2), 233-258.

Frazzini, A., & Pedersen, L. H. (2014). Betting against beta. *Journal of financial economics*, 111(1), 1-25.

Freedman, A. (2022, August 25). *Texas bans BlackRock, UBS, others over ESG investing*. Axios. https://www.axios.com/2022/08/25/texas bans-blackrock-ubs-esg-backlash

Freedman, M., & Jaggi, B. (1986). Pollution performance of firms from pulp and paper industries. *Environmental Management*, 10(3), 359- 365.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), 210- 233. https://doi.org/10.1080/20430795.2015.1118917

Friedman, M. (1970). A theoretical framework for monetary analysis. *journal of Political Economy*, 78(2), 193-238.

Frooman, J. (1997). Socially irresponsible and illegal behavior and shareholder wealth: A meta-analysis of event studies. *Business & society*, 36(3), 221-249.

Godfrey, P. C., Merrill, C. B., & Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic management journal*, 30(4), 425-445.

Griffin, J. J., & Mahon, J. F. (1997). The corporate social performance and corporate financial performance debate: Twenty-five years of incomparable research. *Business & society*, 36(1), 5- 31.

Hart, S. L., & Christensen, C. M. (2002). The great leap: Driving innovation from the base of the pyramid. *MIT Sloan management review*, 44(1), 51.

Hebb, T., Hawley, J. P., Hoepner, A. G., Neher, A. L., & Wood, D. (Eds.). (2016). *The Routledge handbook of responsible investment*. Oxon: Routledge.

Jain, P., Vyas, V., & Roy, A. (2017). Exploring the mediating role of intellectual capital and competitive advantage on the relation between CSR and financial performance in SMEs. *Social Responsibility Journal*.

Kurapatskie, B., & Darnall, N. (2013). Which corporate sustainability activities are associated with greater financial payoffs?. *Business strategy and the environment*, 22(1), 49-61.

Leone, S. (2021). *US Funds dataset from Yahoo Finance*. [Data set]. Kaggle. https://www.kaggle.com/datasets/stefanoleone992/m utual-funds-and-etfs

Lev, H. (2021, December 13). *ESG funds set a new record inflow by doubling in 2021*. ConserviceESG. https://www.gobyinc.com/esg-funds-new-record inflow-2021/

Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative science quarterly*, 48(2), 268-305.

McAbee, J. (2022, June 1). *What is Stakeholder Theory?*. Wrike. https://www.wrike.com/blog/understanding stakeholder-theory/

McGuire, J. B., Schneeweis, T., & Branch, B. (1990). Perceptions of firm quality: A cause or result of firm performance. *Journal of management*, 16(1), 167- 180.

McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate social responsibility and firm financial performance. *Academy of management Journal*, 31(4), 854-872.

McWilliams, A., & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. *Academy of management review*, 26(1), 117-127.

Michelon, G., Boesso, G., & Kumar, K. (2013). Examining the link between strategic corporate social responsibility and company performance: An analysis of the best corporate citizens. *Corporate social responsibility and environmental management*, 20(2), 81-94.

Mishra, S., & Suar, D. (2010). Does corporate social responsibility influence firm performance of Indian companies?. *Journal of business ethics*, 95(4), 571- 601.

Murray, K. B., & Montanari, J. B. (1986). Strategic management of the socially responsible firm: Integrating management and marketing theory. *Academy of management review*, 11(4), 815- 827.

Naylor Association Management Software. (n.d.). *Naylor Association management software*. The Forum for Sustainable and Responsible Investment. Retrieved January 9, 2023, from https://www.ussif.org/about

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta analysis. *Organization studies*, 24(3), 403-441.

Posner, B. Z., & Schmidt, W. H. (1992). Values and the American manager: An update updated. *California Management Review*, 34(3), 80- 94.

Preston, L. E., & O'bannon, D. P. (1997). The corporate social-financial performance relationship: A typology and analysis. *Business & Society*, 36(4), 419-429.

Roman, R. M., Hayibor, S., & Agle, B. R. (1999). The relationship between social and financial performance: Repainting a portrait. *Business & society*, 38(1), 109-125.

Ruf, B. M., Muralidhar, K., Brown, R. M., Janney, J. J., & Paul, K. (2001). An empirical investigation of the relationship between change in corporate social performance and financial performance: A stakeholder theory perspective. *Journal of business ethics*, 32(2), 143-156.

Seifert, B., Morris, S. A., & Bartkus, B. R. (2004). Having, giving, and getting: Slack resources, corporate philanthropy, and firm financial performance. *Business & society,* 43(2), 135-161.

S&P 500. (2020), S&P 500 (^GSPC) [Data Set]. *YahooFinance*, https://finance.yahoo.com/quote/%5EGSPC/history? p=%5EGSPC

Spectrum News Staff. (2022, July 27). *Gov. DeSantis takes aim at 'woke corporations'.* Spectrum News. https://www.baynews9.com/fl/tampa/news/2022/07/2 7/speaking-in-tampa--gov--desantis-takes-aim-at--woke-corporations

Tang, Z., Hull, C. E., & Rothenberg, S. (2012). How corporate social responsibility engagement strategy moderates the CSR–financial performance relationship. *Journal of management Studies*, 49(7), 1274-1303.

Van Beurden, P., & Gössling, T. (2008). The worth of values–a literature review on the relation between corporate social and financial performance. *Journal of business ethics,* 82(2), 407-424.

Van der Laan, G., Van Ees, H., & Van Witteloostuijn, A. (2008). Corporate social and financial performance: An extended stakeholder theory, and empirical test with accounting measures. J*ournal of Business Ethics*, 79(3), 299- 310.

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance–financial performance link. *Strategic management journal*, 18(4), 303-319.

Whelan, T., Atz, U., & Clark, C. (2022). ESG and financial performance: Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015–2020. *NYU, ESG and Performance.*

Yahoo! Finance. (2022, May 25). Parnassus Core Equity Fund - Investor Shares (PRBLX) Stock Price, News, Quote & History - Yahoo Finance. https://finance.yahoo.com/quote/PRBLX?p= PRBLX

8 Appendix

8.1 Cumulative Return

A cumulative return on an investment is the aggregate amount that the investment has gained or lost over time, independent of the amount of time involved. The cumulative return is expressed as a percentage, and it is the raw mathematical return of the following calculation:

((current price of security)-(original price of security)) / original price of security

Cumulative returns in this study were derived from annual total returns.

8.2 Annualized Return

An annualized total return is the geometric average amount of money earned by an investment each year over a given time period. The annualized return formula is calculated as a geometric average to show what an investor would earn over a period of time if the annual return was compounded.

Annualized returns in this study were derived from annual total returns, and calculated using Microsoft Excel.

8.3 Standard Deviation

Standard deviation is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance. The standard deviation is calculated as the square root of variance by determining each data point's deviation relative to the mean.

Standard deviations in this study were derived from annual total returns.

8.4 Downside Deviation

Downside deviation is a measure of downside risk that focuses on returns that fall below a minimum threshold or minimum acceptable return (MAR). It is used in the calculation of the Sortino ratio, a measure of risk-adjusted return.

Downside deviations in this study were derived from annual total returns.

8.5 Sharpe Ratio

The Sharpe ratio compares the return of an investment with its risk. It's a mathematicalexpression of the insight that excess returns over aperiod of time may signify more volatility and risk, rather than investing skill. Formula:

$(R_p - R_f) / \sigma_p$

where R_p = return of portfolio, R_f = risk-free rate, and σ_p = standard deviation of portfolio.

Sharpe ratios in this study were derived from annual total returns and the annualized cumulative return of the rolling 3-month T-bill rate.

8.5.1 Annualized Cumulative Return of the Rolling 3-Month T-Bill Rate

Rolling 3-month T-bill rates for each year from 2005 to 2020 were provided by MarketWatch, linked here: https://www.marketwatch.com/investing/bond/tmub musd03m/download- data?countrycode=bx&-mod=mw_quote_tab.

The annualized cumulative return of the rolling 3-month T-bill rate in this study was derived from rolling 3-month T-bill rates.

8.6 Sortino Ratio

The Sortino ratio is a variation of the Sharpe ratio that differentiates harmful volatility from total overall volatility by using the asset's standard deviation of negative portfolio returns—downside deviation—instead of the total standard deviation of portfolio returns. The Sortino ratio takes an asset or portfolio's return and subtracts the risk-free rate, and then divides that amount by the asset's downside deviation.

Formula:
$(R_p - R_f) / \sigma_p$

where R_p = return of portfolio, R_f = risk-free rate, and σ_d = downside deviationof portfolio.

Sortino ratios in this study were derived from annual total returns and the annualized cumulative return of the rolling 3-month T-bill rate.

8.7 Active Return

Active return is the percentage gain or loss of an investment relative to the investment's benchmark. Active returns in this study were derived from annualized returns.

8.8 Tracking Error

Tracking error is the divergence between the price behavior of a position or a portfolio and the price behavior of a benchmark. Tracking error is reported as a standard deviation percentage difference, which reports the difference between the return an investor receives and that of the benchmark they were attempting to imitate.

Tracking errors in this study were derived from standard deviations.

8.9 Information Ratio

The information ratio (IR) is a measurement of portfolio returns beyond the returns of a benchmark, usually an index, compared to the volatility of those returns.

Formula:
((Portfolio Return) – (Benchmark Return))/Tracking Error

Information ratios in this study were derived from annualized returns and standard deviations.

Berkeley
Economic
Review