Three Essays in Microeconomics

Navin Kumar
Advisors: S. Anukriti, Arthur Lewbel, Richard Tresch, and Utku Unver

A thesis submitted to the faculty of the Department of Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy
Boston College
Morrissey College of Arts and Sciences Graduate School
July 2021

Abstract

In this series of essays, I apply the tools of economics to a variety of real world problems. The first essay looks at the impact of a gun control regulation on mortality and crime. A third of US states have removed all restrictions on carrying concealed handguns. This might decrease crime by invisibly arming law-abiding citizens, or increase it by eliminating penalties for criminals. It could have no effect at all, because handguns are easily hidden, so anyone who wished to carry a gun was already doing so. I compare counties along the borders of states that liberalized concealed carry to contiguous counties in neighboring states that did not, using mortality and crime micro-data. I find that deregulation had no impact on homicide, violent crime, firearm mortality, firearm usage, or firearm ownership.

The second essay, co-authored with Sajala Pandey, looks at the impact of an earthquake in Nepal on child development. Biologists have posited that prenatal maternal stress (PNMS) has an adverse impact on child development, possibly via the process of epigenetic imprinting which occurs during the first trimester. Researchers have attempted to study this link by using natural disasters as a source of exogenous variation. A shortcoming of these studies is that natural disasters may also affect prenatal healthcare provision, either by decreasing it's provision (due to infrastructure being destroyed) or increasing it (thanks to aid flowing into the region.)

We look at the impact of 2015 Earthquake in Nepal on children who were (a) in utero at the time of the earthquake and (b) in areas severely affected by it. Consistent with theories from PNMS, we find that the earthquake adversely impacted their height-for-age, and the effects were concentrated on individuals who were in the first trimester of gestation. These negative effects were entirely offset by an increase in the consumption of antenatal healthcare. We find that the earthquake resulted in a *improvement* in development indicators for those children who were in severely affected areas but not in-utero at the time of the earthquake, highlighting the importance of healthcare provision in early childhood.

The third essay, co-authored with Andrew Copland, proposes a solution to the problem of assigning multiple scarce goods to agents in the absence of prices, for example assigning seats in courses to students in a university. Students submit a list of preferences over courses, a lottery for rankings is held, and an algorithm allocates each student their top available course, reversing their ranks at the end of each round. Then, for each student, the algorithm compares their outcomes to the outcomes generated by every alternative ordering they could have set. Whenever such revisions result in more preferred outcomes, their preferences are replaced with the alternative. Our solution is non-dictatorial and Pareto optimal. When it converges without encountering a loop, it is strategy-proof. It retains properties even in small economies. We compare our algorithm to alternatives.

Acknowledgments

I would like to thank my advisors, S. Anukriti, Arthur Lewbel, Richard Tresch, and Utku Unver for their invaluable advice and guidance.

I would especially like to thank Richard Tresch, who is likely the kindest teacher to have graced the halls of Boston College.

I would like to thank my co-authors, Andrew Copland and Sajala Pandey, whose brilliance is matched only by their patience.

Finally, I would like to thank my brother and my parents, without whose support this project would have been impossible.

Chapter 1. Gun Control and Crime: Evidence from Concealed Carry Laws in the United States

1 Introduction

The United States of America has an unusually high incidence of violent crime - its annual homicide rate is 5.3 per 100,000 residents[1] as opposed to the OECD average of 3.2.[2] Residents often acquire firearms for self-defense.[3] However, the US also experiences 40,000 firearm-related deaths annually. Gun control laws aim to balance the defensive and recreational use of guns with the prevention of violent crime and firearm deaths.

This paper analyzes one such regulation in depth: the unrestricted carry of concealed firearms in public. Proponents of Unrestricted Concealed Carry (UCC) argue that it decreases violent crime, because it is harder for criminals to distinguish between armed and unarmed victims. Opponents argue that UCC increases crime, because criminals themselves can now be legally armed while waiting for an opportunity to commit a crime. If intercepted and searched by the police, they face no repercussions when caught with a concealed weapon. A third possibility is that these two effects offset each other, resulting in no change in crime, but an increase in firearm ownership and usage. A final possibility is that such deregulation will have no effect at all, because the preceding restrictions were difficult to enforce. A person carrying a concealed handgun is unlikely to be stopped and searched by the police, and therefore has little disincentive to carry. Criminals, and even people with no malicious intent, do not need to wait for deregulation to carry a concealed firearm. I call this argument *neutrality*, and find support for it.

At the federal level, there are no laws governing the carry of firearms in public spaces. At the state level, open carry laws govern the act of publicly carrying a weapon[4] "in plain sight", e.g., in a hip or shoulder holster. Concealed carry laws govern the public carry of a weapon that is *not* in plain sight e.g. in one's purse, or glove compartment, or waistband. No issue-regimes ban concealed carry al-

[1]Federal Bureau of Investigation (2016)
[2]OECD 2016.
[3]Pew Research Center 2017.
[4]"Weapon" here includes firearms and knives, but is not limited to them. For example, the state of Florida considers pepper spray to be a weapon.

together. May-issue regimes allow concealed carry with a permit, but give local authorities broad discretion when issuing these permits. Shall-issue regimes do *not* give local authorities any discretion; applicants are entitled to a permit if they meet certain conditions such as background checks, safety classes, safety tests, etc. Unrestricted concealed carry regimes are the most permissive, allowing residents to carry a concealed firearm without a permit. Table 1 presents the concealed carry regimes in the US and maps them to the states that have adopted them.

The existing literature on concealed carry studies the transition of US states from may-issue to shall-issue regimes. In 1960, the vast majority of US states had may-issue regimes in place. In response to rising crime in the 1970s, states started adopting shall-issue regimes[5] and by 2011, 36 states had done so. Lott and Mustard (1997) claimed that, had all states in the US adopted shall-issue laws, "1,570 murders, 4,177 rapes, and over 60,000 aggravate assaults would have been avoided yearly." This claim was challenged by researchers criticizing their data and methodology.[6] In 2005, the National Research Council released a report on gun violence that concluded that there was insufficient evidence to show that shall-issue laws either decreased or increased crime,[7] leading to another round of debate.[8] More recent papers call for caution in analyzing shall-issue laws, finding ambiguous impact.[9]

Unrestricted concealed carry is a relatively new development in gun control. In 2003, Alaska's legislature became the first to adopt it. In 2010, Arizona became the second. As of May 2021, 20 out of 50[10] US states have unrestricted concealed carry regimes in place, and yet little research has been done on the impact of these laws. This paper provides, to the best of my knowledge, the first comprehensive analysis of this development in gun control.

The impact of gun control laws can be difficult to identify because of the endogeneity of gun laws, gun ownership, and crime. Perhaps loose gun control laws lead to more criminals acquiring guns. Perhaps high criminal victimization can

[5]Winkler 2011.

[6]Black and Nagin (1998) argued that their results were sensitive to small changes in the model and sample. Ayres and J. J. Donohue (2003) argued that "more refined" analysis showed the opposite - that crime rose as a result of shall-issue laws. Maltz and Targonski (2002) pointed out they used imputed data to analyze crime, which it ill-suited for. Helland and Tabarrok (2004), using "placebo laws" find that the impact on crime is not well-estimated. These critiques prompted replies from the authors - see Lott (1998), Lott and Whitley (2003), and Plassmann and Whitley (2003)

[7]National Research Council 2005.

[8]See Moody and Marvell (2008), J. Donohue and Ayres (2009), Moody and Marvell (2009), Abhay Aneja, Donohue III, and Zhang (2011), Moody, Lott, and Marvell (2013), J. Donohue, Aneja, and Weber (2017), McElroy and Wang (2017).

[9]Durlauf, Navarro, and Rivers (2016) explicitly argue that "one should be cautious in using the results from any particular model to inform policy decisions." Gius (2018) found that different empirical methods yielded different conclusions about the link between shall-issue laws and murder rates, reversing Gius (2014). Manski and Pepper (2018) concludes that "[shall-issue] laws increase some crimes, decrease other crimes, and have effects that vary over time for others." Gresenz (2018) reviews the literature and concludes that evidence "that shall-issue concealed-carry laws may increase violent crime is limited. Evidence for the effect of shall-issue laws on total homicides, firearm homicides, robberies, assaults, and rapes is inconclusive."

[10]Figure 1 shows the liberalization of concealed carry in the US.

lead to voters demanding looser gun laws to protect themselves. Perhaps a culture of gun ownership leads to looser laws as well as more crime. Perhaps it's some combination of these factors. It can be difficult to extract causation from correlation.

I tackle endogeneity using a paired border counties difference-in-difference approach. Counties along the borders of US states are similar to contiguous counties in neighboring states. Each county along the border of a treatment state is matched to an adjacent county in a control state, creating a pair. If a county has more than one neighbor in the control state, more than one pair of counties is created. For outcomes of interest, I construct a variable measuring the difference between counties within a pair, and then track this variable after the implementation of UCC. To this end, I deploy an OLS regression that includes a full set of county fixed effects, county-trends, and state-month fixed effects.[11] To the best of my knowledge, this is the first paper in the gun control literature to compare border counties to identify the impact of a firearm regulation.[12]

Data on mortality are drawn from the Multiple Cause of Death (MCOD) files, maintained by the Centers for Disease Control. These files contain the universe of US death certificates and are made available to researchers upon approval. This micro-data allows me to analyze the impact of UCC with a degree of granularity that previous researchers did not have access to.

Homicide is one of the most important outcomes of interest, and the one most robust to manipulation by law enforcement agencies. I find that UCC had no impact on homicide. In my preferred specification, there is a statistically insignificant change in homicide, corresponding to an increase of only 0.15%. The insignificance and magnitude of this coefficient is inconsistent with the theories that UCC increased or decreased violent crime, and consistent with the theory that it had no impact.

I supplement the analysis of mortality data with data on crime. These data come from the incident-level files which are collated by the National Incident Based Reporting System (NIBRS), which is maintained by the Federal Bureau of Investigation (FBI). Law enforcement organizations use their own record management systems to send incident reports to the FBI. The NIBRS includes incident-level data on crimes committed, and these data include the date, time, type of incident, and weapon used, among other factors. Consistent with neutrality, I find that UCC has no detectable impact on homicide, and no impact on sexual assault, aggravated assault, or robbery.[13]

[11]State-month fixed effects are used to control for seasonal variation

[12]A possible shortcoming of the border counties approach is that the relatively small number of border counties and the relatively short time-span that this policy has been in existence for may lead to insufficient sample size for analysis. It is therefore supplemented - at each step - with a traditional all-county difference-in-difference approach, with the increased sample size compensating for the lack of plausible identification. I use an OLS regression with a full set of county fixed effects, county-trends, and region-month fixed effects. The results from this analysis are, overall, consistent with the results from the border counties analysis.

[13]Over the course of this paper, I will be using the term "no impact" to describe estimations where the coefficient

It is possible that the policy led to an increase in criminals *and* law abiding citizens carrying weapons, resulting in no overall change in crime, but nonetheless increasing gun ownership and usage. Following Niekamp (2018), I use the number of armed arrestees as a proxy for day-to-day gun usage, using arrestee reports from the NIBRS. Following Moody and Marvell (2005), I use the number of firearm suicides as a proxy for gun ownership, using mortality data from the MCOD. I find that UCC has no impact on gun usage or gun ownership, supporting the theory that UCC is truly neutral.

To test the possibility that changes in the legal system that confounded this analysis, I construct a placebo variable of crimes that are not plausibly affected by gun laws. To test the possibility that UCC led criminals to switch to less confrontational crimes, I check its effect on burglary, theft, and simple assault. I find that UCC has no impact on the placebo variable or non-violent crimes.

These conclusions must be interpreted carefully. This paper does *not* claim that concealed carry itself is harmless or completely lacking in impact. All states studied in this paper transitioned from a shall-issue regime to an unrestricted regime. It is possible that *this* transition is harmless while the transition from may-issue to shall-issue is harmful (or beneficial, or harmless.) Finally, it is possible that while this crime has no effect in the short-to-medium run, it may have effects in the long run, though a synthetic control analysis of a state that adopted this policy early finds no support for this theory.

These findings have two major implications for policy. First, Unrestricted Concealed Carry does not decrease crime, and policy makers will have to look elsewhere for ways to do so. Second, it does not increase gun violence, so gun control activists would not see substantial gains from revoking such legislations.

2 Context

In 1976, Washington DC banned the purchase of handguns by residents. In 2008, the Supreme Court of the United States ruled this to be a violation of the Second Amendment of the US Constitution, asserting that it guaranteed the right of citizens to possess firearms.[14] However, the ruling gave states the right to impose "reasonable restrictions", such as banning convicted felons from possessing a firearm. Most importantly, for the purpose of this project, the ruling explicitly did not guarantee the right of citizens to carry a concealed firearm. Supreme Court Justice Antonin Scalia wrote:

was statistically insignificant *and* fulfilled one or both of two conditions: either the data had sufficient sample size to rule out minimum detectable effects exceeding 10%, or the estimated effect sizes were less than 10%. If the estimates were statistically insignificant but did not fulfill either of these conditions, I use the term "no detectable impact" instead. See the section on empirical strategy for more.

[14] Supreme Court of the United States 2008.

> Like most rights, the Second Amendment right is not unlimited. It is not a right to keep and carry any weapon whatsoever in any manner whatsoever and for whatever purpose: For example, *concealed weapons prohibitions have been upheld under the Amendment or state analogues* [Emphasis added]. The Court's opinion should not be taken to cast doubt on longstanding prohibitions on the possession of firearms by felons and the mentally ill, or laws forbidding the carrying of firearms in sensitive places such as schools and government buildings, or laws imposing conditions and qualifications on the commercial sale of arms.

There are currently no federal laws governing the possession a weapon in public - all such laws are made by local or state legislatures.

The concealed carry of a weapon refers to the practice of carrying a weapon, such as a handgun or a knife, in a public place in such a way that it is not "in plain sight". It could be on one's person (e.g. in a purse or one's waistband) or in close proximity (e.g. in a glove compartment.) While the exact definition of "plain sight" varies from jurisdiction to jurisdiction, there are, broadly speaking, 4 types of concealed carry regimes:

- *No Issue:* Residents are not allowed to carry a concealed handgun in a public place. American Samoa is one such place.
- *May-Issue:* These are jurisdictions that allow the concealed carry of a handgun, but local authorities have discretion over whether permits are granted or not. Many counties and cities in may-issue regimes are no-issue regimes in practice, as law enforcement agencies are often reluctant to issue such permits.[15]
- *Shall-Issue:* These are jurisdictions that require a permit to carry a concealed gun, and local authorities have little or no discretion over who gets a permit. Authorities are required to issue a permit if the applicant has fulfilled conditions determined by the law. These conditions can include background checks, minimum ages, safety and training classes, safety tests, and proficiency tests, amongst other things. The laws that create such regimes are referred to as "Right-To-Carry" laws or "Shall-Issue" laws.
- *Unrestricted Concealed Carry*: A permit is not needed to carry a concealed handgun. This term is a slight misnomer - there are still restrictions on who can carry a concealed weapon (e.g. people with felony convictions and underage people are not allowed to) and where (e.g. guns are banned in courthouses and schools.)

[15]Cramer and Kopel 1995.

Table 1 presents the 4 categories from least-to-most permissive and enumerates the states that fall under these policies.

In 1813, Kentucky and Louisiana banned concealed carry, and were the first states to do so. Over the course of the nineteenth century, an increasing number of US states followed suit and by the middle of the twentieth century, most states enacted concealed carry restrictions.[16] A notable exception is Vermont. In 1903, the city of Rutland in Vermont attempted to regulate the carry of weapons in public spaces. However, the Vermont Supreme court ruled it a violation of the state constitution.[17] For this reason UCC is also known as Vermont carry.

In 1961, Washington adopted a shall-issue concealed carry permitting regime. In 1980, responding to a sharp increase in crime, Indiana became the second US state to do so[18]. This process accelerated in the 1980s and 1990s, as an increasing number of states transitioned from no-issue or may-issue regimes to shall-issue regimes. By 2011, 36 states had enacted such laws. The bulk of the empirical literature on concealed carry examines the transition from may-issue regimes to shall-issue regimes[19]

In 2003, Alaska became the first state to switch from a shall-issue regime to an unrestricted concealed regime. In 2010 Arkansas became the second. This process accelerated in the 2010s and, at the time of publication, 20 out of 50 US states have UCC regimes in place. Figure 1 shows the change in policy in US States over time, while Table 2 shows the dates at which unrestricted concealed carry went into effect in various states. Every treatment state in the sample is a state that transitioned from shall-issue to unrestricted concealed carry.

3 Conceptual Framework

I follow the economic tradition established by Becker (1968), treating the criminal as an agent weighing the costs and benefits of committing a crime. The benefit is the material or immaterial gain from successfully committing a crime. The costs are many; among others, there is the guilt or shame from having committed a crime, the opportunity cost of committing a crime, the social costs that arise if ones community discovers ones wrongdoing, the legal penalties that arise if one is caught and convicted, and the physical danger from a target who resists. It is that last two of these that deregulating concealed carry affects, in opposite directions.

Criminals who confront their victims, such as muggers, have to consider whether their potential victims are armed or not. Visibly armed targets are easy to avoid. Under no-issue or may-issue concealed carry regimes, targets are likely unarmed.

[16]Winkler 2011.
[17]Supreme Court of Vermont 1903.
[18]Winkler 2011.
[19]Gresenz 2018.

Under shall-issue regimes, targets are more likely to be invisibly armed, but statutory and bureaucratic hurdles may keep the probability low. The likelihood of an encounter with an invisibly armed victim is highest in unrestricted concealed carry regimes, and this may discourage criminals who fear for their health or safety.

However, now criminals themselves may be armed without fear of consequence. Criminals motivated by material gain are often opportunistic, waiting for the right moment and target to enact a crime.[20] A mugger who openly carries a weapon can be avoided by potential victims. In a shall-issue concealed carry regime, a mugger who applies for a permit leaves a paper trail, which they may wish to avoid. A mugger who carries a concealed weapon without a permit risks being intercepted by law enforcement officials who can arrest them under any regime other than unrestricted concealed carry. Under unrestricted concealed carry, criminals without a felony record can carry a concealed weapon without fear of being arrested by law enforcement. Thus potential legal penalties are reduced.

It is also possible that the law leads to both negative and positive effects, with both criminals and law-abiding citizens acquiring firearms. Indeed, it's possible that these effects could cancel out, leading to the incorrect conclusion that the law hd no effect, simply because crime remained unchanged. The law's effects will be seen in the increased incidence of gun ownership and gun usage, though perhaps not the crime data.[21]

Neutrality can arise under UCC if potential criminals (or law-abiding citizens) are unlikely to be stopped and searched by police officers, making the cost of covertly carrying a weapon negligible. A change in the costs of carrying a weapon in this situation would therefore have no effect.[22]

4 Data

4.1 Data on Mortality

When a US resident dies, their death is registered using the US Standard Certificate of Death which is filled out by a medical examiner. The causes of death are classified according to the International Classification of Diseases (10th Revision). These death certificates are collected as part of the National Vital Statistics Systems (NVSS), an inter-governmental system for the sharing of data on births and deaths in the US. The NVSS is the product of coordination between state health departments

[20] Felson and Clarke 1998.

[21] Previous versions of this paper included a formal mathematical model of concealed carry and violence. This model was not be estimated; its purpose was solely to illustrate the incentives facing criminals and non-criminals under different carry regimes, and was excluded for the purpose of brevity.

[22] It is important to note that such laws may have different effects on different sub-populations, such as women, African-Americans, or the residents of rural and urban areas. The results for such groups is similar to results for the overall paper, and they were excluded in the interests of brevity.

and the National Center for Health Statistics, a division of the Centers for Disease Control and Prevention.

Since 2005, granular death certificate data has not made publicly available due to concerns about privacy. However, approved researchers are given access to the Multiple Cause of Death (MCOD) files, which contain anonymized data on each death that occurs in the US. The observational unit in these data is a death. These data include the date and time of death, the county of residence, the county of death, race, sex, age, the cause of death, and so on.

For this paper, I use data from 2008 to 2015, the last year for which these data were made available. The year 2008 was chosen as a cut off to be consistent with the analysis for crime data. For all US counties, I construct a panel of population-adjusted monthly deaths caused by firearm homicide, firearm suicide, firearm accidents, and firearm deaths where the intent is unknown. The number of non-firearm homicides is also used, as a placebo. To remain consistent with the crime analysis, these data are aggregated as deaths per 100,000 residents instead of the per million residents that is standard in the public health literature. Table 5 shows the control and treatment states in the all county sample and the number of counties in each state that are part of the panel. Table 6 shows the treatment states that are in the border county sample, the number of pairs in them, the control states with whom they share a border, and the number of pairs of border counties the treatment and control states have. For example, Arizona has a total of 22 border county pairs. It's border with California generates 5 pairs, it's border with Colorado generates 1 pair and so on. Note that each county can be part of more than one pair.

Table 7 shows summary statistics for mortality in the pre-treatment year 2010, and also compares the level of firearm mortality in the treatment and control counties. For the all county sample I use a two-sample t-test, and for the border county sample I use a paired t-test. For both samples the differences between the level of firearm homicide in control and treatment counties are statistically insignificant.

Suicides committed with a firearm are used as a proxy for firearm ownership. This follows the literature established by researchers such as Moody and Marvell (2005), Cook and Ludwig (2006), and Kovandzic, Schaffer, and Kleck (2013). For the border county sample, the differences between firearm suicides in the control and treatment counties are statistically insignificant. I observe a statistically significant difference in the level of firearm suicides in the all-county sample, suggesting that the treatment states tend to have higher levels of gun ownership than the control states. This is one of the reasons I eschew a spatial regression discontinuity approach in favor of a difference-in-difference approach.

4.2 Data on Crime

Data on crime come from the National Incident Based Reporting System (NIBRS), which collects incident-level data on violent (and some non-violent) crimes that occur in a law-enforcement agency's jurisdiction. Law enforcement agencies generate NIBRS incident data via their own record management systems, which relay the data to the Federal Bureau of Investigation who collate the data publicly accessible in collaboration with the Inter-University Consortium for Political and Social Research. The data include the incident date, time, offenses committed, whether these offenses were completed or attempted, the type of location (restaurant, residence etc), and the type of weapon used. Of particular importance to this project is that the data records the demographic characteristics of the both the victim and offender - including age, ethnicity, gender - and what relationship (if any) they had to each other.[23]

Crimes in a county are tabulated by the number of incidents of the crime per month, per 100,000 residents. The crimes that the FBI categorizes as violent crimes - homicide, sexual assault, aggravated assault, and robberies - are also the crimes most frequently studied in the literature on gun control.[24] I also look at burglaries (also known as breaking and entering) and thefts. A placebo crime variable is constructed by adding up the number of incidents of crimes that are plausibly unaffected by UCC laws. These include, and are limited to, blackmail, counterfeiting, fraud, embezzlement, pornography, gambling, sex work, bribery, bad checks, loitering, vagrancy, and driving under the influence.

The NIBRS data is tabulated from 2008-2016. 2016 is the last year for which the data was available at the time of my analysis. The further back one goes, the fewer law enforcement agencies participate in the NIBRS, and so I chose 2008 as an (unavoidably arbitrary) cutoff, and later will check that the entry of agencies had no impact on the conclusions of my analysis. Table 4 shows the states that are in the border-county sample, and Table 3 shows the states that are in the all-county sample. Figure 2 maps the counties that are in our border county sample. The NIBRS also includes data on the arrests made by law enforcement officers. These data include the demographic characteristics of the arrestee, a cross-reference to the crime committed, and information about whether or not the arrestee was armed and, if so, with what. This is true for all arrests, not just those made for violent or

[23]It should be noted that the researchers studying shall-issue concealed carry (SICC) regimes did not have access to incident-level data, because relatively few police departments were part of the NIBRS system at the time states transitioned into SICC regimes. Researchers relied on the Summary Reporting System in which agencies reported their crime statistics to state agencies, which then conveyed them to the FBI. Maltz and Targonski (2002) point out that these data are not appropriate for county-level analysis, since the FBI often imputed data when it was not available. Aggregating incident data myself using the NIBRS allows me to avoid this problem. Furthermore, I can aggregate the data at the level of the month, an option not available to researchers of the shall-issue laws, who relied on annual data.

[24]See, for example, the first paper on concealed carry written by an economist, Lott and Mustard (1997)

gun related crimes; a person arrested for driving under the influence of alcohol will be recorded as being armed if they are carrying a handgun, even if that handgun was legal.

A county-month panel dataset was constructed with 15,600 observations for 182 pairs of border counties spread across 5 treatment states and 14 control states, tracking the number of incidents of crime (per 100,000 residents) for the time period 2008-16. Due to limitations on the availability of data, this does not include all possible border county pairs. Figure 2 shows the counties that are included in the paired border county sample. For the all-county method, a monthly panel data was constructed for 1740 counties spread over 5 treatment states and 35 control states for the period 2008-2016. When deploying the border county method, the data has a total of 160,000 observations.

I use the number of arrests as a proxy for day-to-day usage of guns. In this, I am following Niekamp (2018), who found that gun usage sharply rose at the beginning of hunting seasons in US states using this measure. Since the population of people being arrested is likely dissimilar to the general population, this should not be though of as a precise measure of the general level of day-to-day gun usage, but only as a proxy useful for measuring changes in usage.

4.3 Miscellaneous Data

Data on Law Enforcement Organizations come from the Law Enforcement Management and Administrative Statistics (LEMAS) survey. These data are used to check the similarity of law enforcement in treatment counties and their paired control counties. They include responsibilities, expenses, salaries and special pay, the demographic characteristics of their officers, and so on. This analysis can be found in the section on empirical strategy. The survey is conducted irregularly by the Department of Justice. I use the 2013 survey, which is in the pre-treatment period for all treatment counties, except those in Arkansas.

Data on county characteristics are from the US Census Bureau's Statistical Compendia program, and are used to check the similarity of border counties along demographic characteristics. While the program is now defunct, it was active in the pre-treatment period of this paper's analysis. The fraction of adults with at least a high school education is used as a measure of education. The fraction of voters voting for the Republican party (in the 2008 election) is used as a measure of voting patterns. Other measures include the headcount ratio to measure poverty, median individual income as a measure of income, people per square mile as a measure of density, and the unemployment rate. This analysis can also be found in the section on empirical strategy.

5 Empirical Strategy

5.1 Border County Difference-in-Difference

For the crime data, I construct a county-by-month panel dataset for the time period 2008-16, match neighboring border counties to each other, and then construct a variable

$$Y'_{csc's'my} = Y_{csmy} - Y_{c's'my}$$

where Y_{csmy} is the outcome Y for county c in treatment state s in month m and year y. $Y_{c's'my}$ is the outcome Y for adjacent county c' in control state s' such that $s \neq s'$ and s' has never had concealed carry over this period.

I then estimate

$$Y'_{csc's'my} = \alpha + \beta D_{smy} + \lambda_{csc's'} + \lambda_{my} + \lambda_{sm} + u_{csmy} \tag{1}$$

where $Y'_{csc's'my}$ is the difference between the outcome of interest in a county c in treatment state s and c' in control state s' where c' is contiguous to c. D_{smy} is a policy indicator that takes the value 1 if the policy is active in state s in month-year my. $\lambda_{csc's'}$ is a full set of county pair fixed effects, while λ_{my} is a full set of time period fixed effects. λ_{sm} is a set of state-specific month-wise fixed effects, meant to capture seasonal effects. Note that each pair of counties is treated as if it were a separate geographical unit, so if a county c in state s neighbors two contiguous counties c^1 and c^2 in state s', they will be a part of two border pairs: csc^1s' and csc^2s'. Standard errors are clustered by treatment state and use wild cluster bootstrapping to estimate confidence intervals.

The coefficient on the policy dummy is interpreted as the change in crime in the treatment county relative to it's neighbor. A positive coefficient indicates an increase in crime or mortality, while a negative coefficient indicates a decrease in the same.

The crime data has 15,663 units spread over 5 treatment states and 182 pairs. Table 4 shows the treatment and control states in the crime sample, and the number of border county pairs corresponding to each state. 2016 is the last year for which data is available. 2008 precedes the earliest instance of legalization by three years. Participation in the NIBRS has risen over time, an implication of which is that the further one goes back in the dataset, the fewer agencies are a part of it. I judged 2008 to be a reasonable cut-off point.

The mortality data is a sample of 488 border county pairs spread over 12 treatment states and 26 control states, for the years 2008-2015[25], for a total of approximately 52,500 observations. Table 6 shows the treatment and control states in the

[25] Data for the years 2016-17 have been applied for and will be included in future versions of this paper

mortality sample and the number of border county pairs corresponding to each state.

5.2 All County Difference-in-Difference

For the NIBRS data, I construct a county-by-month panel dataset for the time period 2008-16, and then estimate

$$Y_{csrmy} = \alpha + \beta D_{smy} + \lambda_{cs} + \lambda_{my} + \lambda_{rm} + u_{csrmy} \tag{2}$$

where Y_{csrmy} is the variable of interest in county c, state s, climatic region r, month m, and year y. D_{smy} is a binary which takes the value 1 if state s had adopted UCC before month m and year y and is 0 otherwise. λ_{cs} is a full set of county fixed effects, λ_{my} is a full set of month-year fixed effects, and λ_{rm} is a set of region-specific month-wise fixed-effects to capture seasonal variation. Standard errors are clustered by state.

We are interested in the coefficient on D_{smy}. A negative coefficient indicates that crime has fallen in treatment states relative to control states, while a positive coefficient would indicate the opposite. The size of the coefficients estimates the change in crime or mortality per 100,000 residents.

Table 3 shows the states in the crime data. There is a total of 160,511 observations spread over 40 states for the years 2008-2016.

Table 5 shows the states in the mortality data. I construct a county-month panel dataset for 3,102 counties spread over 12 treatment states and 39 control states for the years 2008-2015, for a total of approximately 372,000 observations.

With a slight abuse of terminology, I will refer to the estimations of Equation 1 as an analysis of the "border county sample" and the estimations of Equation 2 as an analysis of the "all county sample."

5.3 Validity

5.3.1 Parallel Trends

It is important that trends in mortality and crime before the implementation of a policy are similar between control and treatment groups in the border counties sample. To that end, I estimate the following equation:

$$Y'_{csc's'my} = \alpha + \sum_{j=-24}^{24} \beta_j D_{smy}(my = k + j) + \lambda_{csc's'} + \lambda_{my} + \lambda_{sm} + u_{csc's'my} \tag{3}$$

where $Y'_{csc's'my}$ is the difference between outcomes in a county c in treatment state s and its neighbor c' in control state s' during month m of year y. Here, k is the month and year at which a policy actually goes into effect. $D_{smy}(my = k + j)$ is a set of policy indicators that take the value 1 if the UCC has been adopted in state s in month m and year y with lags and leads. There are 24 policy indicators that lead the treatment effect, and 24 indicators that lag behind it. $\lambda_{csc's'}$ is a full set of county fixed effects, while λ_{my} is a full set of time period fixed effects. λ_{sm} is a full set of state-month fixed effects meant to control for seasonal variation in crime. For the parallel trends assumption to be considered valid, $\beta_j = 0$ for all $j < 0$. If the impact of the law were neutral, $\beta_j = 0$ for all $j \geq 0$ as well.

For the all county sample, I estimate

$$Y_{csrmy} = \alpha + \sum_{j=-24}^{24} \beta_j D_{smy}(my = k + j) + \lambda_{cs} + \lambda_{my} + \lambda_{rm} + u_{csrmy} \tag{4}$$

where Y_{csrmy} is the variable of interest in county c in state s in region r for month m in year y. k is the month and year at which a policy actually goes into effect. $D_{smy}(my = k + j)$ is a set of policy indicators that take the value 1 if the UCC has been adopted in state s in month m and year y with lags and leads. There are 24 policy indicators that lead the treatment effect, 24 policy indicators that lag behind it. λ_{cs} is a full set of county fixed effects, while λ_{my} is a full set of time period fixed effects. λ_{rm} is a full set of region-month fixed effects meant to control for seasonal variation in crime. For the parallel trends assumption to be considered valid, $\beta_j = 0$ for all $j < 0$. If the impact of the law were neutral, $\beta_j = 0$ for all $j \geq 0$ as well.

The parallel trends assumption holds for the NIBRS crime data. Figure 3 plots the β_j coefficients of border county Equation 3, and Figure 4 plots the β_j coefficients of all counties Equation 4. For homicide, rape, aggravated assault, and robbery, these coefficients are close to zero from month to month, and are statistically insignificant.

The parallel trends assumption also holds for the MCOD mortality data. Figure 5 plots the β_j coefficients of border county Equation 3, and Figure 6 plots the β_j coefficients of all counties Equation 4.

5.3.2 Balance

I check that the panel dataset is balanced along various dimensions by comparing pre-treatment levels of demographic characteristics, geographical characteristics, crime, and mortality in the treatment and control counties. The paired two-sided t-test is used to perform this comparison for the border county sample, and a regular two-sided t-test for the all-county sample.

Table 10 compares the county characteristics of control and treatment counties in the border county sample in 2010, before the implementation of UCC. It shows the results of a paired t-test done on various measures of interests. All figures are normalized for population. The treatment and control counties have similar levels of sworn and non-sworn personnel, full-time and part-time personnel, similar operating budgets, and equipment. Furthermore, treatment and control counties have similar high school graduation rates, Republican vote shares, unemployment, poverty, and population density. They differ in following respects: median income is lower in treatment counties (38k vs 40k), and the racial composition skews more white (89% vs 85%).

Data from the NIBRS is used to average monthly incidence of crimes (normalized for population) between treatment and control counties. Table 9 shows that treatment and control counties have similar pre-treatment levels of homicide, sexual assault, and robbery, though treatment counties have significantly more aggravated assault. For this reason, as well as the differences in racial composition and income levels, I eschew a spatial discontinuity approach in favor of a difference-in-difference approach that requires similar pre-treatment trends, but not levels, in the outcomes of interest.

Table 7 compares 2010 mortality in the control and treatment counties. In the all county sample, there are similar levels of pre-treatment homicide, but significantly more firearm suicides, suggesting that the treatment states have similar levels of crime, but higher levels of initial gun ownership. However, this problem does not exist in the border counties sample, where control and treatment counties show similar initial levels of both firearm homicides and firearm suicides, suggesting that the border county panel is better balanced than the all county panel.

5.3.3 Sample Size

As most of the resulting coefficients are statistically insignificant, I try to show that there I have well-identified zero instead of a sample size that is too small to detect an effect. As a way to intuitively communicate the adequacy of the sample size of a specification, I follow Bloom (1995) and report the Minimum Detectable Effect (MDE) for each estimate, setting the $\alpha = 0.05$ and $\beta = 0.2$. If the MDE is (say) 10% that means that we have sufficient sample size to rule out a change of 10% or more i.e. we can rule out a coefficient of 11% but not 9%. The MDE will sometimes be larger than the coefficient and sometimes be smaller; neither of these outcomes have any particular meaning.

The Minimum Detectable Effects are discussed using the following terminology from Bloom (1995): 0-10% are small effects, 10-25% are modest effects, and over 25% are large effects. Ideally, the sample should have sufficient size to rule out even small effects, but that is impossible, so we focus on ruling out modest effects instead.

In discussing my results, whenever the coefficient of the estimates is statistically insignificant, I use the phrase "no detectable impact." I use the stronger phrase "no impact" in situations where either the sample size is high enough or the effect size is low enough to warrant such language. The thresholds for "high enough" and “low enough” are unavoidably arbitrary. Since we are focusing on ensuring that the we have sufficient sample size to rule out modest effects, I use this phrase if the MDE is 10% or less. To remain consistent with this, I also use the phrase if the size of the effect is 10% or less..

6 Results

6.1 Impact on Homicide

Homicide is an important measure for many reasons. First, it is a serious crime, and carries steep legal penalties. Second, it is important to the discourse on firearm regulation; gun rights activists and gun control activists discuss it at length, though they have differing beliefs about the direction of the impact of legislation. Third, it is a reliable indicator of the level of crime in a jurisdiction. Police departments sometimes attempt to improve their statistics by downplaying serious crimes e.g. reclassifying an incident of aggravated assault as simple assault. It is difficult to do this for homicide.

The most complete source of data on homicide is the mortality data drawn from the universe of death certificates. My preferred specification is Equation 1, which analyzes the sample of paired border counties, which are demographically very similar to each other. Panel A of Table 11 presents the coefficient estimates. Consistent with the theory that the law had no impact, the increase in the incidence of firearm homicide is 0.15% and statistically insignificant.

Panel B shows the estimates of Equation 2, which is a difference-in-difference analysis of firearm mortality for all US counties. It shows a statistically insignificant 4.8% increase in firearm homicides in the all-county sample, consistent with UCC having no impact.

Perhaps the results are being confounded by an overall change in criminal law, rather than merely gun laws? To check that this is not the case, I apply the above methodology to non-firearm homicides. The coefficient estimates show a statistically insignificant decrease of 4.9%, consistent with UCC having no impact.[26]

Turning to the crime data, Panel C and Panel D of Table 12 shows the estimates of Equation 2 - the all country difference-in-difference analysis - for violent crimes. Column 1 shows the coefficients of the policy dummy for this equation. Unrestricted concealed carry had little to no effect on homicides. Looking at all homicides, there is a statistically insignificant coefficient of -0.015, corresponding to a decline of 6.2%. Results for firearm homicides are starker: the coefficient is -0.003, corresponding to a decline of only 2.3%. These estimates mirror the mortality

[26] In addition to this, I checked the legislative records of states that adopted these policies and was unable to find any major changes to the legal system. This method is, admittedly, somewhat harder to replicate.

estimates.

The weakest supporting evidence can be seen in Panel A and Panel B of Table 12, which shows the estimates of Equation 1 - the border county difference-in-difference approach - for violent crimes. Column 1 shows the coefficients of the policy dummy for homicide. From this estimation, we may conclude that UCC has no detectable impact on homicide, though we cannot go as far as to assert that it had no impact, because the coefficient and MDE are both quite large.

6.2 Impact on Other Violent Crimes

Next, I consider the impact of deregulating concealed carry on violent crimes that do not result in a fatality. My conclusions are consistent with the results for homicide - deregulating concealed carry does not seem to change the incidence of violent crimes such as sexual assault, aggravated assault, or robbery. This is true even if the analysis were to be restricted to crimes committed with a firearm.

Sexual assault is analyzed in Column 2 of Table 12. With sufficient sample size to rule out modest effects, I find that UCC had no detectable impact on sexual assault - the coefficient on the incidence of sexual assault is negative and statistically insignificant in both the border county and all county samples.

Aggravated assault refers to either a physical altercation in which serious injuries are sustained, or in which a weapon is brandished. Column 3 of Table 12 analyzes the impact of UCC on aggravated assault. With sufficient sample size to rule out modest effects, I find that UCC had no impact on aggravated assault, which is consistent with neutrality.

Robberies are analyzed in Column 4 of Table 12. These refer to crimes that involved confrontation between an assailant and victim; incidents that didn't are classified as thefts. Column 4 of Table 12 analyzes the impact of UCC on robberies. I find that that UCC had no detectable impact on robberies. I have sufficient sample size to rule out modest effects in the all-county panel, and large effects in the border county panel, and so would not assert that these estimations support neutrality, merely that they do not contradict it.

6.3 Impact on Nonviolent Crimes

A concern in the literature on law and economics is that criminals may respond to increased penalties for crime A by switching to crime B, leaving the overall level of crime unchanged. Consider, for example, burglary. A criminal may prefer to be a burglar rather than a mugger if their potential victims are completely disarmed. If their victims can have a firearm at home but not in public, the criminal may opt to be a mugger instead, since their victims are likely to be disarmed. If concealed carry in public is liberalized as well, then burglary becomes relatively more attractive again.

This does not appear to be the case for concealed carry. Table 13 shows the impact of UCC on burglary and theft. The coefficient for the policy dummy is negative and statistically insignificant for each crime in the border county sample as well as the all county sample. For each crime, there is sufficient sample size to rule out modest changes in the incidence of these crimes, consistent with UCC having no impact on the crime.

7 Is the Policy Neutral, or Are Effects Offsetting Each Other?

It is possible that the adoption of UCC would lead to both criminals and non-criminals acquiring firearms, and the positive and negative effects of the policy cancel each other out. If this is true, there should be an increase in ownership and usage, even in the absence of an increase or decrease in crime. In this section, I test the possibility that effects are offsetting each other by looking at the impact of UCC on gun ownership rates and day-to-day gun usage.

Gun suicides are a commonly used proxy for gun ownership, because people who wish to kill themselves and own one or more firearms are likely to use a gun for that purpose. Column 3 of Table 11 shows the impact of the legalization of UCC on firearm suicide in the mortality data. Consistent with neutrality, the coefficient corresponds to a statistically insignificant 2.5% decline in firearm suicides in the border counties sample. It corresponds to a statistically insignificant 9% decline in the all-county sample, which is within our definition of "no impact" but on the border of it.

Police arrest reports note whether arrestees were armed at the time of arrest, even if they were not being arrested for a violent or firearm-related crime. I use this variable - constructed with the NIBRS data - as a proxy for day-to-day gun usage. Table 14 shows the impact of UCC on the number of armed arrestees. The coefficient estimates are statistically insignificant and have opposite signs, with the border county estimate showing a 10% decrease and the all county estimate showing a 5% increase. In both cases, there is insufficient sample size to rule out these modest effect sizes.

Overall, the data on suicides and arrests do not support the theory that UCC led to an increase in gun ownership and usage by criminals and law abiding citizens alike, with these effects canceling each other out. This, in turn, lends weight to the theory that neither group was waiting on the state government to deregulate concealed carry.

8 Other Robustness tests

8.1 The Entry of Law Enforcement Organizations into the Data

The NIBRS is a relatively recent program, and has expanded considerably over the period of time covered by this analysis. The expansion allows me to conduct analysis that previous researchers were not able to do, but carries a drawback: if agencies that joined late had relatively fewer incidents than the average agency that could bias the coefficient estimates downwards. Alternatively, if such agencies had more incidents than the average, it could bias the coefficient estimates upwards.

To test this possibility, I restrict the analysis to agencies that were present for the entirety of the period 2008-16. These results are presented in Table 15. I find no detectable impact of UCC on violent crimes in the border counties sample, though the sample size has been diminished to the point that I cannot assert that there was no impact. In the all county sample, I find that UCC had no impact on violent crime.

8.2 Spillovers

It is possible that criminals in border counties, now faced with invisibly armed residents, will travel to neighboring counties without UCC regimes to commit crimes. Thus, crime would rise in the neighboring county, leading to erroneous conclusions being drawn from the analysis, which focuses on the difference between crime in treatment and control counties. Specifically, the coefficient estimates will be biased downwards.

To test this possibility, I estimate

$$Y_{c's't} = \alpha + \beta D_{st} + \lambda_{c's'} + \lambda_t + u_{c's't}$$

where $Y_{c's't}$ is the outcome of interest in control county c' in state s'. D_{st} is a policy indicator that takes the value 1 if the policy is active in state s at time period t. $\lambda_{c's'}$ is a full set of county fixed effects, while λ_t is a full set of time period fixed effects. As before, standard errors are estimated using the wild cluster bootstrapping method.

Table 16 shows the results of this analysis. The adoption of UCC in treatment states had no detectable impact on trends in crime in the control counties that border them i.e. there are no detectable changes in crime that arise due to the implementation of this policy.

8.3 Placebo Tests

It is possible that such laws are adopted along with other changes to policing or the judicial system - ones that negate the non-zero impact of this law - resulting in misleading conclusions being drawn from difference in difference analysis. To test this theory I construct a placebo variable consisting of crimes that are plausibly unaffected by gun control measures, including and limited to blackmail, counterfeiting, fraud, embezzlement, pornography, gambling, sex work, bribery, bad checks, loitering, vagrancy, and driving under the influence. The impact of legalizing unrestricted concealed carry on this placebo variable is then analyzed using the same tools as before.

Table 13 shows the impact of UCC on this placebo variable. In Panel A - an estimation using the border county method - I find that UCC had no impact on placebo crimes, with a coefficient whose magnitude corresponds to a 4.6% statistically insignificant increase in crime.

In the all county sample, there is a statistically significant 13% decline, which is a reminder that the states that adopted such laws are different from states that did not, and why focusing on border counties is so important.

8.4 Time after implementation

A shortcoming of the data is that most of the states that deregulated concealed carry did so quite recently, creating a relatively brief post-treatment window. It is possible that the law does not have an immediate impact but does have some sort of long term impact.

An exception to this is Arkansas, which legalized unrestricted concealed carry in 2013 and therefore has a relatively long post-treatment period. However, it is but one state, and it would be unwise to draw any strong conclusions from it. To address this problem, I follow Abadie, Diamond, and Hainmueller (2010) in applying the synthetic control method to this problem. I construct a state by applying weights to a control group of states such that the resultant "state" is similar to Arkansas along various demographic dimensions. I then check that there was no divergence between these two "states" after the implementation of unrestricted concealed carry.

After excluding the other treatment states in the data, a synthetic control for Arkansas was created using pre-intervention state-level demographic data to assign weights to the remaining US states, including the fraction of Whites, median income, high school graduation rates, Republican vote share, the unemployment rate, and the poverty rate. The analysis was performed using the `synth` program in STATA.

Figure 7 shows the impact of the unrestricted concealed carry in Arkansas vs it's synthetic control. Before and after the implementation of the policy, trends in both "states" closely matched each other, which supports neutrality.

9 Discussion & Conclusion

Proponents of unrestricted concealed carry argue that it will decrease crime as it would lead to more law-abiding citizens carrying firearms. Opponents argue that it will increase crime, as it would lead to more criminals carry firearms. This paper finds no support for either of these theories, but does find support for the theory that it had no effect at all. In my preferred specification, I find that UCC leads to a statistically insignificant 0.15% increase in homicide, which is consistent with no impact. Additionally, I find that it had no impact on sexual assault, robbery, aggravated assault, or non-violent crimes.

Of course, this alone does not imply that the law had no impact at all - it could have led to criminals and law-abiding citizens carrying more firearms but these two effects cancelling each other out. However, the data also shows that UCC has no impact on gun ownership, using suicides as a proxy, or on day-to-day usage, using arrestee reports as a proxy. These findings contradict the "offsetting" theory. This leaves us with only one explanation: that the law did not influence people's decision to carry concealed firearms, and so had no effect - positive or negative - on crime.

Why *didn't* the law have any effect? A concealed firearm - by definition - is difficult to spot, and it is unlikely that a person would be stopped and searched. There is little reason for someone to *not* be armed should they so wish, whether they had malicious intent or not. This paper analyzes the transition from permit-based concealed carry to unrestricted concealed carry. The effect of such laws comes not from a previously forbidden action becoming legal, but from a change to costs. It is possible that the costs of acquiring a permit were not prohibitive, so any law-abiding citizen who wanted to carry a firearm did so. Criminals likely banked on not being stopped and searched before they committed their crime and continued their previous behavior.

None of this is to say that concealed carry is harmless or completely lacking in impact. It is possible that the transition from shall-issue to UCC is harmless while the transition from may-issue to shall-issue is harmful (or beneficial, or harmless.) Furthermore, the synthetic control analysis of Arkansas notwithstanding, it is possible that the policy may lead to long term effects that are yet to be seen. Nonetheless, the neutrality of UCC is consistent with the recent literature or shall-issue that finds that the impact of this transition to be ambiguous.

There are three takeaways from this analysis.

First, UCC is not a successful crime-fighting tool. This is unfortunate, as it requires no taxation or public funding, and would have been a cost-effective tool had it succeeded. Policy-makers who wish to reduce the incidence of violent crime will need to look elsewhere.

Second, UCC is not a dangerous form of deregulation. This is fortunate, as a third of US states have transitioned into UCC regimes and it is comforting to know that the residents of these states have not been subject to increased violence as a result.

Third, the enforceability of laws should be a major concern while drafting gun control legislation or prioritizing gun control activism. Laws governing the carry of weapons may be impractical, and so activists may wish to turn their attention to other measures such as universal background checks, or legislation that ensures that people with domestic violence records are restricted from purchasing firearms via the NICS. Siegel et al. (2019) provides a overview of the effect of gun control measures affect mortality, and is an excellent resource for gun control activists.

This paper hits many firsts in the analysis of gun control policy. It is the first to use a border county difference-in-difference approach to identify the impact of such a policy. It is the first to use granular death certificate data. It is also one of the few to comprehensively analyze this relatively new gun control policy, as an increasing number of states adopt it. An important contribution of this paper is methodological - to highlight how large datasets and the tools available to modern economists and criminologists can be used to answer important questions related to public policy, especially important and difficult to answer questions such as the impact of a specific gun control policies.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". In: *Journal of the American Statistical Association* 105.490, pp. 493–505.

Aneja, Abhay, John J Donohue III, and Alexandria Zhang (2011). "The impact of right-to-carry laws and the NRC report: lessons for the empirical evaluation of law and policy". In: *American Law and Economics Review* 13.2, pp. 565–631.

Ayres, Ian and John J Donohue (2003). "Shooting down the more guns, less crime hypothesis". In: *Stanford Law Review* 55, pp. 1193–1312.

Becker, Gary S. (1968). "Crime and Punishment: An Economic Approach". In: *Journal of Political Economy* 76.2, pp. 169–217. DOI: 10.1086/259394.

Black, Dan A and Daniel S Nagin (1998). "Do right-to-carry laws deter violent crime?" In: *The Journal of Legal Studies* 27.1, pp. 209–219.

Bloom, Howard S (1995). "Minimum detectable effects: A simple way to report the statistical power of experimental designs". In: *Evaluation review* 19.5, pp. 547–556.

Cook, Philip J and Jens Ludwig (2006). "The social costs of gun ownership". In: *Journal of Public Economics* 90.1-2, pp. 379–391.

Cramer, Clayton E and David B Kopel (1995). "Shall Issue: The New Wave of Concealed Handgun Permit Laws". In: *Tennessee Law Review*.

Donohue, John, A Aneja, and KD Weber (2017). *Right-to-Carry Laws and Violent Crime: Assessment Using Panel Data and a State-Level Synthetic Controls Analysis*.

Donohue, John and Ian Ayres (2009). "More guns, less crime fails again: the latest evidence from 1977–2006". In: *Econ Journal Watch*.

Durlauf, Steven N, Salvador Navarro, and David A Rivers (2016). "Model uncertainty and the effect of shall-issue right-to-carry laws on crime". In: *European Economic Review* 81, pp. 32–67.

Federal Bureau of Investigation (2016). "Crime in the United States". In:

Felson, Marcus and Ronald V Clarke (1998). "Opportunity Makes the thief". In: *Police Research Series* 98, pp. 1–36.

Gius, Mark (2014). "An examination of the effects of concealed weapons laws and assault weapons bans on state-level murder rates". In: *Applied Economics Letters* 21.4, pp. 265–267.

— (2018). "Using the Synthetic Control Method to Determine the Effects of Concealed Carry Laws on State-Level Murder Rates". In: *International Review of Law and Economics*.

Gresenz, Carole Roan (2018). "Effects of Concealed-Carry Laws on Violent Crime". In: *RAND Corporation*.

Helland, Eric and Alexander Tabarrok (2004). "Using placebo laws to test "more guns, less crime"". In: *Advances in Economic Analysis & Policy* 4.1.

Kovandzic, Tomislav, Mark E Schaffer, and Gary Kleck (2013). "Estimating the causal effect of gun prevalence on homicide rates: A local average treatment effect approach". In: *Journal of Quantitative Criminology* 29.4, pp. 477–541.

Lott, John R (1998). "The concealed-handgun debate". In: *The Journal of Legal Studies* 27.1, pp. 221–243.

Lott, John R and David B Mustard (1997). "Crime, deterrence, and right-to-carry concealed handguns". In: *The Journal of Legal Studies* 26.1, pp. 1–68.

Lott, John R and John Whitley (2003). “Measurement error in county-level UCR data”. In: *Journal of Quantitative Criminology* 19.2, pp. 185–198.

Maltz, Michael D and Joseph Targonski (2002). “A note on the use of county-level UCR data”. In: *Journal of Quantitative Criminology* 18.3, pp. 297–318.

Manski, Charles F and John V Pepper (2018). “How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions”. In: *Review of Economics and Statistics* 100.2, pp. 232–244.

McElroy, Marjorie B and Will Wang (2017). “Seemingly Inextricable Dynamic Differences: The Case of Concealed Gun Permit, Violent Crime and State Panel Data”. In:

Moody, Carlisle E, John R Lott, and Thomas B Marvell (2013). “Did John Lott provide bad data to the NRC? A note on Aneja, Donohue, and Zhang”. In: *Econ Journal Watch* 10.1.

Moody, Carlisle E and Thomas B Marvell (2005). “Guns and crime”. In: *Southern Economic Journal*, pp. 720–736.

— (2008). “The debate on shall-issue laws”. In: *Econ Journal Watch* 5.3, p. 269.

— (2009). “The debate on shall issue laws, continued”. In: *Econ Journal Watch* 6.2, p. 203.

National Research Council (2005). *Firearms and violence: A critical review*. National Academies Press.

Niekamp, Paul (2018). “Good Bang for the Buck: Effects of Rural Gun Use on Crime”. In:

OECD (2016). “OECD Better Life Index”. In:

Pew Research Center (2017). “America’s Complex Relationship with Guns”. In:

Plassmann, Florenz and John Whitley (2003). “Confirming “more guns, less crime"”. In: *Stanford Law Review*, pp. 1313–1369.

Siegel, Michael et al. (2019). “The Impact of State Firearm Laws on Homicide and Suicide Deaths in the USA, 1991–2016: a Panel Study”. In: *Journal of General Internal Medicine*, pp. 1–8.

Supreme Court of the United States (2008). “District of Columbia v. Heller”. In: *554 U.S. 570*.

Supreme Court of Vermont (1903). “State v. Rosenthal”. In: *75 Vt. 295, 55 A. 610*.

Winkler, Adam (2011). *Gunfight: The battle over the right to bear arms in America*. WW Norton & Company.

Appendix: Figures & Tables

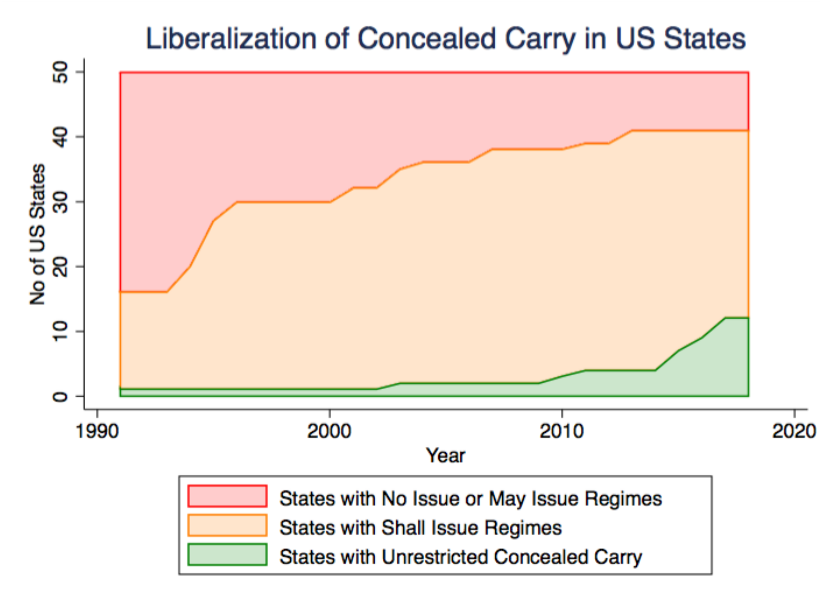

FIGURE 1: TRENDS IN CONCEALED CARRY

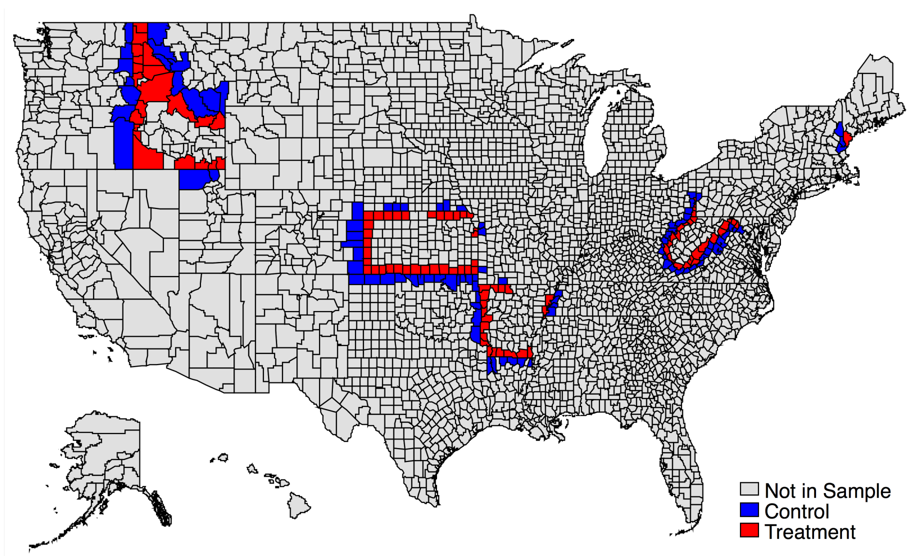

FIGURE 2: COUNTIES IN BORDER COUNTIES SAMPLE

Data is from the National Incident Based Reporting System, maintained by the Federal Bureau of Investigation

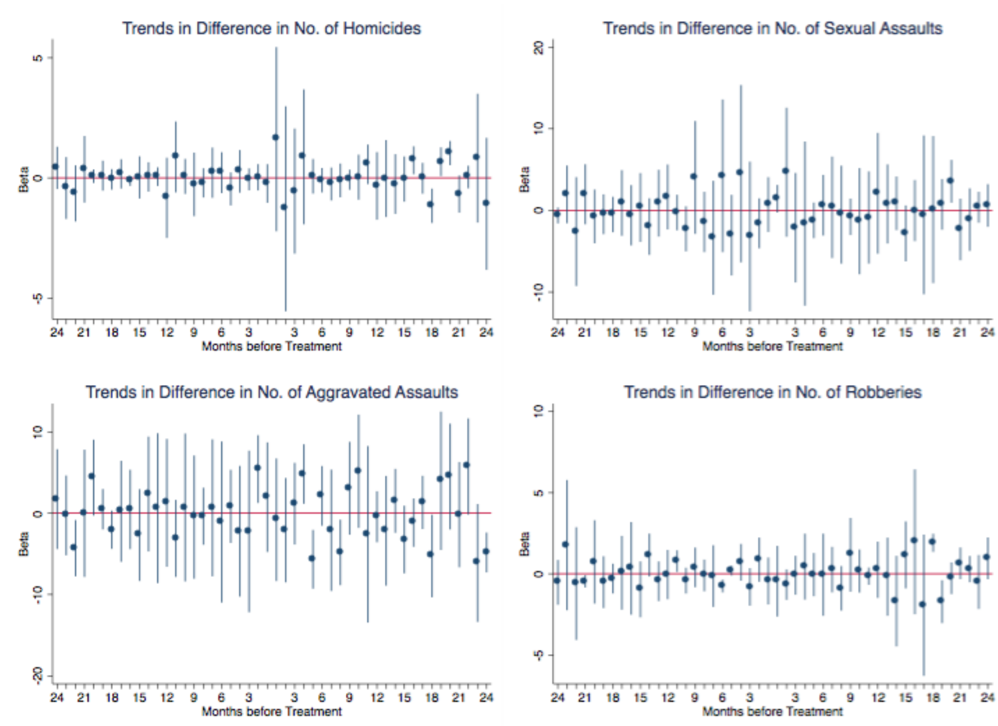

FIGURE 3: PRETRENDS IN CRIME FOR BORDER COUNTIES

Data is from the National Incident Based Reporting System, maintained by the Federal Bureau of Investigation

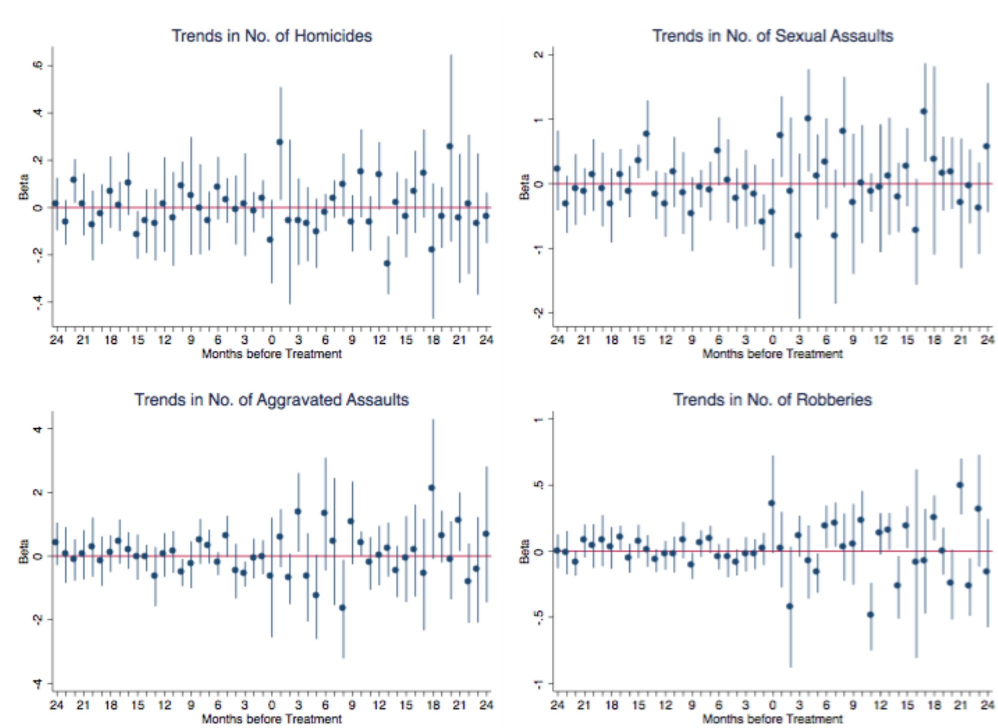

FIGURE 4: PRETRENDS IN CRIME FOR ALL COUNTIES

Data is from the National Incident Based Reporting System, maintained by the Federal Bureau of Investigation

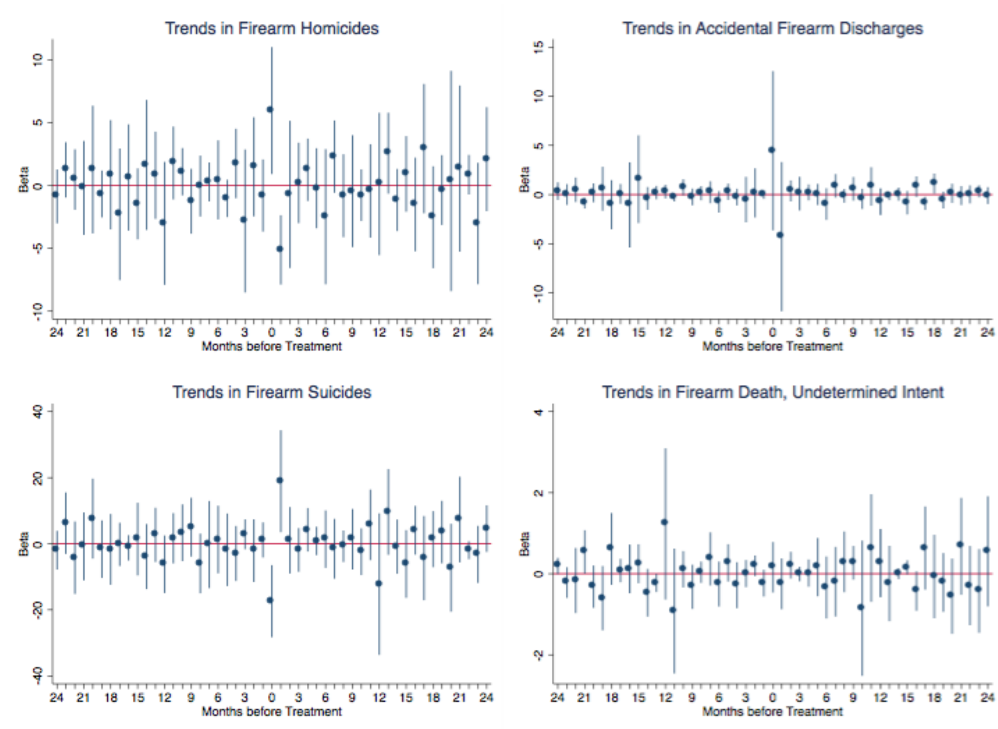

FIGURE 5: PRETRENDS IN MORTALITY FOR BORDER COUNTES

Data is from the Multiple Cause of Death Files

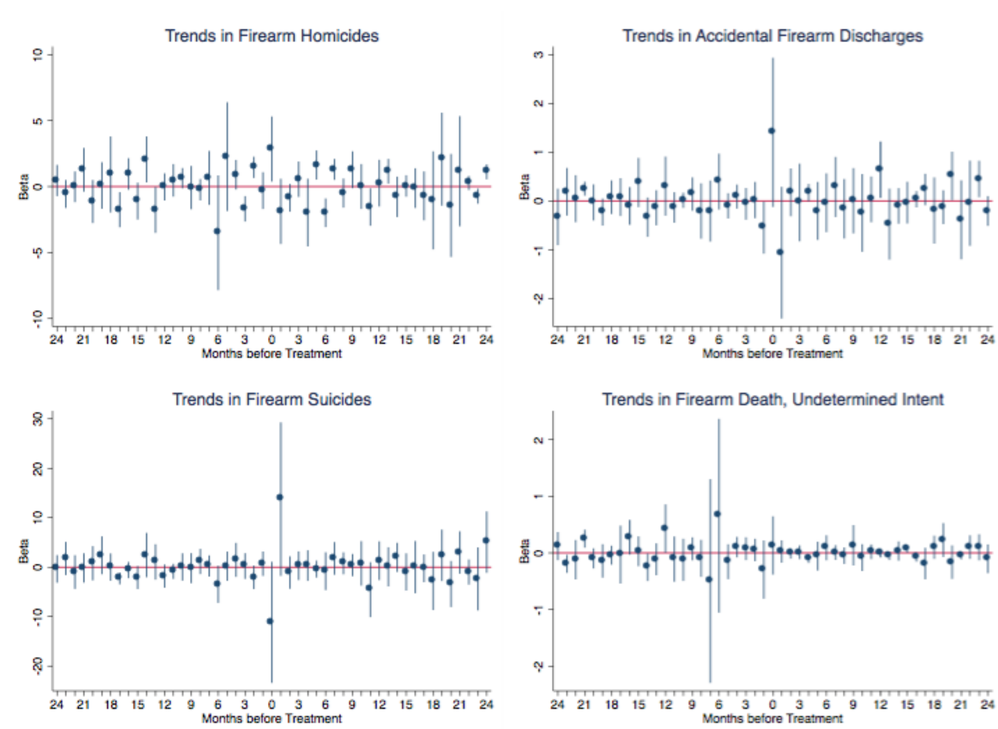

FIGURE 6: PRETRENDS IN MORTALITY FOR ALL COUNTIES

Data is from the Multiple Cause of Death Files

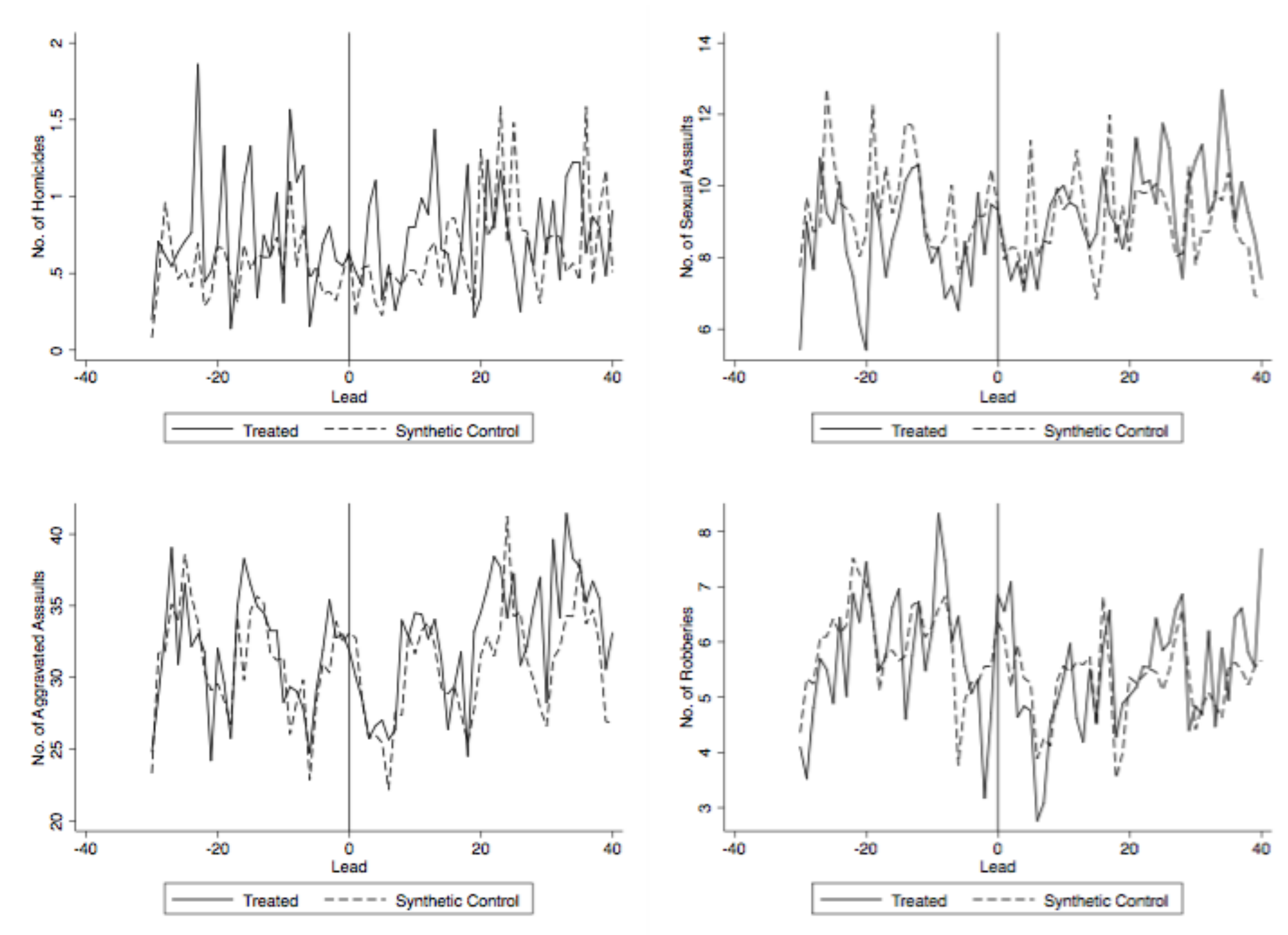

Figure 7: Arkansas vs Synthetic Control

Appendix: Tables

Name	Allows Concealed Carry?	Mandates Issue of Permit?	States
No Issue	No	No	American Samoa, Northern Mariana Islands
May Issue	Yes	No	CA, CT, DE, HI, MA, MD, NJ, NY, PR, RI
Shall Issue	Yes	Yes	AL, CO, DC, FL, GA, Guam, IL, IN,LA, MI, MN, MN, NE, NV, NM, NC, OH, OR, PN, SC, TX, VA, WA, WI
Unrestricted	Yes	N/A	AK, AR, AZ, IA, ID, KS, KY, ME, MO, MS, ND, NH, OK, SD, TN, UT, VT, WV, WY

Table 1: Concealed Carry Regimes in the USA, from Least to Most Permissive

State	Date effective	Note
Vermont	n/a	Ruled a constitutional right by the Vermont State Supreme Court in 1903
Alaska	September 9, 2003	
Arizona	July 1, 2011	
Wyoming	July 1, 2011	
Arkansas	August 16, 2013	
Idaho	July 1, 2015	Legalized outside cities
	July 1, 2016	Legalized within cities
Kansas	July 1, 2015	
Mississippi	July 1, 2015	
Maine	July 1, 2015	
West Virginia	May 24, 2015	
Missouri	January 1, 2017	
New Hampshire	February 22, 2017	
North Dakota	August 1, 2017	
Kentucky	June 26, 2019	
South Dakota	July 1, 2019	
Oklahoma	November 1, 2019	
Utah	May 5, 2021	
Tennessee	July 1, 2021	
Iowa	July 1, 2021	

Table 2: Dates for the Legalization of Unrestricted Concealed Carry

Control		Treatment	
Statename	Counties	Statename	Counties
Alabama	1	Arizona	2
Colorado	38	Arkansas	69
Connecticut	8	Idaho	35
Delaware	3	Kansas	66
District Of Columbia	1	Maine	4
Illinois	1	West Virginia	49
Iowa	71		.
Kentucky	15		.
Louisiana	13		.
Massachusetts	13		.
Michigan	82		.
Montana	31		.
Nebraska	26		.
New Hampshire	10		.
North Dakota	21		.
Ohio	76		.
Oklahoma	43		.
Oregon	13		.
Rhode Island	5		.
South Carolina	46		.
South Dakota	27		.
Tennessee	94		.
Texas	23		.
Utah	12		.
Virginia	131		.
Washington	7		.
Wisconsin	17		.
Total	828	*Total*	225

Notes: Crime incidence data drawn from National Incident Based Reporting System files, maintained by the Federal Bureau of Investigation. Data sources were limited to municipal police department and county sheriff offices; state and federal agencies were excluded from analysis.

Table 3: States in All-County Crime Sample

Treatment	Total	Control	Pairs
Arkansas	32	Louisiana	11
		Missouri	3
		Oklahoma	12
		Tennessee	6
Idaho	38	Montana	15
		Oregon	9
		Utah	5
		Washington	9
Kansas	56	Colorado	12
		Missouri	5
		Nebraska	13
		Oklahoma	26
Maine	3	New Hampshire	3
West Virginia	53	Kentucky	5
		Ohio	20
		Pennsylvania	1
		Virginia	27
Total	182		.

Notes: Crime incidence data drawn from National Incident Based Reporting System files, maintained by the Federal Bureau of Investigation. Data sources were limited to municipal police department and county sheriff offices; state and federal agencies were excluded from analysis.

TABLE 4: STATES IN BORDER COUNTY CRIME SAMPLE

Control		Treatment	
Statename	Counties	Statename	Counties
Alabama	67	Alaska	25
California	58	Arizona	15
Colorado	62	Arkansas	74
Connecticut	8	Idaho	44
Delaware	2	Kansas	102
District Of Columbia	1	Maine	15
Florida	67	Mississippi	81
Georgia	158	Missouri	115
Hawaii	3	New Hampshire	9
Illinois	102	North Dakota	48
Indiana	91	Vermont	13
Iowa	98	West Virginia	54
Kentucky	119	Wyoming	23
Louisiana	63		.
Maryland	25		.
Massachusetts	13		.
Michigan	82		.
Minnesota	86		.
Montana	52		.
Nebraska	79		.
Nevada	18		.
New Jersey	21		.
New Mexico	32		.
New York	62		.
North Carolina	99		.
Ohio	88		.
Oklahoma	76		.
Oregon	35		.
Pennsylvania	67		.
Rhode Island	4		.
South Carolina	46		.
South Dakota	64		.
Tennessee	94		.
Texas	244		.
Utah	28		.
Virginia	132		.
Washington	38		.
Wisconsin	71		.
Total	2455	*Total*	618

Notes: Mortality data drawn from National Vital Statistics System's Multiple Cause of Death files.

Table 5: States in All-County Mortality Sample

Treatment	Total	Control	Pairs
Arizona	22	California	5
		Colorado	1
		Nevada	2
		New Mexico	8
		Utah	6
Arkansas	32	Louisiana	14
		Oklahoma	9
		Tennessee	6
		Texas	3
Idaho	42	Montana	15
		Nevada	4
		Oregon	7
		Utah	7
		Washington	9
Kansas	61	Colorado	11
		Nebraska	26
		Oklahoma	24
Mississippi	54	Alabama	21
		Louisiana	23
		Tennessee	10
Missouri	64	Illinois	27
		Iowa	18
		Kentucky	5
		Montana	4
		Nebraska	4
		Oklahoma	3
		Tennessee	3
New Hampshire	6	Massachusetts	6
North Dakota	35	Minnesota	11
		Montana	8
		South Dakota	16
Vermont	11	Massachusetts	3
		New York	8
West Virginia	71	Kentucky	5
		Maryland	9
		Ohio	20
		Pennsylvania	10
		Virginia	27
Wyoming	37	Colorado	8
		Montana	11
		Nebraska	6
		South Dakota	7
		Utah	5
Total	435		.

Notes: Mortality data drawn from National Vital Statistics System's Multiple Cause of Death files.

Table 6: States in Border County Mortality Sample

	Treatment Mean	Treatment Obs	Control Mean	Control Obs	Difference
Border Counties (Paired t-test)					
Firearm Homicides	1.925	5,244	1.969	5,244	-0.045
	(0.150)		(0.228)		(0.270)
Accidental Firearm Discharges	0.282	5,244	0.466	5,244	-0.184
	(0.056)		(0.111)		(0.125)
Firearm Suicides	10.147	5,244	9.404	5,244	0.743
	(0.556)		(0.508)		(0.752)
Firearm Death, Undetermined Intent	0.059	5,244	0.017	5,244	0.042*
	(0.024)		(0.007)		(0.025)
Homicide by Unspecified Means	1.593	5,244	0.822	5,244	0.771***
	(0.184)		(0.080)		(0.197)
All Counties (Two sample t-test)					
Firearm Homicides	1.980	7,512	1.874	29,688	0.106
	(0.132)		(0.067)		(0.148)
Accidental Firearm Discharges	0.388	7,512	0.327	29,688	0.061
	(0.074)		(0.035)		(0.079)
Firearm Suicides	9.519	7,512	7.729	29,688	1.790***
	(0.436)		(0.167)		(0.398)
Firearm Death, Undetermined Intent	0.098	7,512	0.088	29,688	0.010
	(0.032)		(0.014)		(0.032)
Homicide by Unspecified Means	1.114	7,512	0.977	29,688	0.137
	(0.112)		(0.043)		(0.103)

Notes: Crime data drawn from National Incident Based Reporting System, maintained by the Federal Bureau of Investigation. Mortality data drawn from Multiple Cause of Death files, maintained by the Centers for Disease Control. All statistics are for year 2010. The p-values assigned to the difference in the all county sample are drawn from a two-sample ttest. The p-values assigned to the difference in border county sample are drawn from a paired ttest. All figures are per 100,000 deaths. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 7: Comparison of Mortality in Treatment and Control Counties, Pre-Treatment

	Treatment	Control	Difference	N
Law Enforcement Characteristics				
No. of Full-Time Sworn Personnel	37.853	28.792	9.061	182
	(5.351)	(5.037)	(7.682)	
No. of Unpaid Sworn Reserve Officers	6.818	3.649	3.169*	182
	(1.592)	(0.903)	(1.897)	
No. of Paid Full-Time Nonsworn Personnel	12.892	13.738	-0.846	182
	(2.070)	(3.116)	(3.865)	
No. of Paid Part-Time Nonsworn Personnel	1.755	1.862	-0.107	182
	(0.446)	(0.459)	(0.624)	
Total Operating Budget (in millions)	3.408	3.133	0.275	182
	(0.493)	(0.590)	(0.812)	
No. of Marked Vehicles Operated	21.406	17.241	4.165	182
	(3.026)	(2.865)	(4.334)	
No. of Unmarked Vehicles Operated	7.255	7.439	-0.184	182
	(1.010)	(1.474)	(1.872)	
Pre-treatment County Characteristics				
High School Graduation Rate in County	83.073	83.562	-0.489	182
	(0.481)	(0.497)	(0.437)	
Percent of Votes Cast for Republicans	63.812	62.629	1.184	182
	(0.844)	(0.877)	(0.767)	
Unemployment Rate in County	8.724	8.426	0.298	182
	(0.236)	(0.235)	(0.194)	
Percent of Population Below Poverty	15.104	15.715	-0.611	182
	(0.416)	(0.450)	(0.440)	
Population per Sq Mile	82.482	88.877	-6.395	182
	(11.947)	(13.873)	(11.579)	
Median Income of County	38468.077	40806.852	-2338.775***	182
	(530.799)	(732.314)	(666.551)	
Percentage of County that is White	88.587	85.341	3.246***	182
	(0.893)	(1.070)	(0.746)	

Notes: Law Enforcement Data is drawn from the 2013 Law Enforcement Management and Administrative Statistics (LEMAS) survey. County characteristic data drawn from the US Census Bureau's Statistical Compendia program. Test of equivalence is paired t-test. Figures are per 100,000 residents. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

TABLE 8: COMPARISON OF TREATMENT AND CONTROL COUNTIES, PRE-TREATMENT

	Treatment		Control		Difference
	Mean	Obs	Mean	Obs	
Border Counties (Paired t-test)					
No. of Homicides	0.118	134	0.154	134	-0.035
	(0.051)		(0.073)		(0.084)
No. of Sexual Assaults	5.303	134	4.704	134	0.599
	(0.523)		(0.553)		(0.754)
No. of Aggravated Assaults	12.515	134	6.068	134	6.447***
	(1.479)		(0.932)		(1.310)
No. of Robberies	2.084	134	1.442	134	0.642
	(0.407)		(0.374)		(0.467)
All Counties (Two sample t-test)					
No. of Homicides	0.197	295	0.255	1,125	-0.058
	(0.060)		(0.041)		(0.085)
No. of Sexual Assaults	5.470	295	6.098	1,125	-0.628
	(0.402)		(0.340)		(0.695)
No. of Aggravated Assaults	11.625	295	8.963	1,125	2.662***
	(0.836)		(0.369)		(0.837)
No. of Robberies	1.462	295	2.336	1,125	-0.874***
	(0.210)		(0.142)		(0.297)

Notes: Crime data drawn from National Incident Based Reporting System, maintained by the Federal Bureau of Investigation. Mortality data drawn from Multiple Cause of Death files, maintained by the Centers for Disease Control. All statistics are for year 2010. The p-values assigned to the difference in the all county sample are drawn from a two-sample ttest. The p-values assigned to the difference in border county sample are drawn from a paired ttest. Firgures are per 100,000 residents. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 9: Comparison of Crime in Treatment and Control Counties, Pre-Treatment

	Treatment	Control	Difference	N
Law Enforcement Characteristics				
No. of Full-Time Sworn Personnel	37.853	28.792	9.061	182
	(5.351)	(5.037)	(7.682)	
No. of Unpaid Sworn Reserve Officers	6.818	3.649	3.169*	182
	(1.592)	(0.903)	(1.897)	
No. of Paid Full-Time Nonsworn Personnel	12.892	13.738	-0.846	182
	(2.070)	(3.116)	(3.865)	
No. of Paid Part-Time Nonsworn Personnel	1.755	1.862	-0.107	182
	(0.446)	(0.459)	(0.624)	
Total Operating Budget (in millions)	3.408	3.133	0.275	182
	(0.493)	(0.590)	(0.812)	
No. of Marked Vehicles Operated	21.406	17.241	4.165	182
	(3.026)	(2.865)	(4.334)	
No. of Unmarked Vehicles Operated	7.255	7.439	-0.184	182
	(1.010)	(1.474)	(1.872)	
Pre-treatment County Characteristics				
High School Graduation Rate in County	83.073	83.562	-0.489	182
	(0.481)	(0.497)	(0.437)	
Percent of Votes Cast for Republicans	63.812	62.629	1.184	182
	(0.844)	(0.877)	(0.767)	
Unemployment Rate in County	8.724	8.426	0.298	182
	(0.236)	(0.235)	(0.194)	
Percent of Population Below Poverty	15.104	15.715	-0.611	182
	(0.416)	(0.450)	(0.440)	
Population per Sq Mile	82.482	88.877	-6.395	182
	(11.947)	(13.873)	(11.579)	
Median Income of County	38468.077	40806.852	-2338.775***	182
	(530.799)	(732.314)	(666.551)	
Percentage of County that is White	88.587	85.341	3.246***	182
	(0.893)	(1.070)	(0.746)	

Notes: Law Enforcement Data is drawn from the 2013 Law Enforcement Management and Administrative Statistics (LEMAS) survey. County characteristic data drawn from the US Census Bureau's Statistical Compendia program. Test of equivalence is paired t-test. Figures are per 100,000 residents. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 10: Comparison of Treatment and Control Counties, Pre-Treatment

	Firearm Mortality				Non-Firearm Mortality
	Homicide	Accident	Suicide	Unk Intent	Other Homicide
	Coef/CI	Coef/CI	Coef/CI	Coef/CI	Coef/CI
Panel A: Border Counties					
Post Treatment	0.003	0.678	-0.254	-0.022	-0.078
	(-0.778,0.783)	(-0.058,1.413)	(-1.776,1.268)	(-0.257,0.213)	(-0.596,0.439)
Observations	52551	52551	52551	52551	52551
Avg Trt Mort	1.925	.282	10.147	.059	1.593
MDE (%)	17.97	46.19	12.67	94.95	26.69
No. of Clusters	12	12	12	12	12
Panel B: All Counties					
Post Treatment	0.096	0.130	-0.862	-0.009	-0.231**
	(-0.486,0.677)	(-0.129,0.389)	(-3.240,1.515)	(-0.081,0.062)	(-0.459,-0.003)
Observations	371965	371965	371965	371965	371965
Avg Trt Mort	1.98	.388	9.519	.098	1.114
MDE (%)	16.67	46.14	11.45	82.34	25.22
No. of Clusters	51	51	51	51	51

P-Values estimated via wildcluster bootstrapping. Data is from the National Vital Statistics System's Multiple Cause of Death Files. Estimates are presented per month, per 100,000 residents. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 11: Impact of Unrestricted Concealed Carry on Mortality

	Homicide Coef/CI	Sx. Assault Coef/CI	Ag. Assault Coef/CI	Robbery Coef/CI
Panel A: Violent Crimes, Border Counties				
Post Treatment	-0.166	-0.796	-4.222	0.424
	[-0.249,-0.082]	[-1.871,0.279]	[-6.892,-1.552]	[-0.092,0.940]
Observations	15663	15663	15663	15663
Mean in Trt Cnty	.315	5.308	14.017	2.112
MDE (%)	50.11	8.71	8.06	12.93
No. of Clusters	5	5	5	5
Panel B: Firearm Crimes, Border Counties				
Post Treatment	-0.049	-0.077*	0.804	0.377
	[-0.235,0.137]	[-0.149,-0.005]	[-0.083,1.691]	[-0.134,0.888]
Observations	15663	15663	15663	15663
Mean in Trt Cnty	.206	.042	2.619	.762
MDE (%)	55.75	67.37	16.23	17.97
No. of Clusters	5	5	5	5
Panel C: Violent Crimes, All Counties				
Post Intervention	-0.015	-0.559	-0.752	0.014
	[-0.019,0.049]	[-1.609,0.491]	[-2.452,0.948]	[-0.222,0.25]
Observations	160511	160511	160511	160511
Mean in Trt Cnty	.242	5.438	12.529	1.485
MDE (%)	23.91	5.9	4.36	9.6
No. of Clusters	40	40	40	40
Panel D: Firearm Crimes, All Counties				
Post Intervention	-0.003	-0.008	0.281	0.067
	[-0.017,0.011]	[-0.018,0.002]	[-0.451,1.013]	[-0.031,0.165]
Observations	160511	160511	160511	160511
Mean in Trt Cnty	.131	.028	2.19	.534
Max ES	32.97	83.41	8.35	14.24
No. of Clusters	40	40	40	40

Notes: Confidence intervals are reported at the 95% level. Confidence intervals for Panels C and D obtained via wildcluster bootstrapping. Data is from the National Vital Statistics System's Multiple Cause of Death Files. All estimates include a year fixed effect to capture overall trends, and month fixed effects to capture seasonal variation. Estimates are presented per month, per 100,000 residents $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 12: Impact of Unrestricted Concealed Carry on Violent Crimes

	Burglary Coef/CI	Theft Coef/CI	Simple Assault Coef/CI	Placebo Coef/CI
Panel A: Border Counties				
Post Treatment	-12.534	-16.048*	-0.855	0.882
	[-19.523,-5.545]	[-22.900,-9.197]	[-8.698,6.988]	[-4.757,6.521]
Observations	15663	15663	15663	15663
Mean in Trt Cnty	40.622	106.013	47.554	19.135
MDE (%)	6.23	5.11	4.59	6.21
No. of Clusters	5	5	5	5
Panel B: All Counties				
Post Intervention	-5.256	-8.482	-2.804	-2.592**
	[-12.888,2.376]	[-15.541,-1.423]	[-6.996,1.388]	[-3.787,-1.397]
Observations	160511	160511	160511	160511
Mean in Trt Cnty	40.84	110.014	46.775	21.757
Mean in Con Cnty	35.408	110.2	46.471	20.059
Power	1	1	1	1
Clustered At	State	State	State	State
No. of Clusters	40	40	40	40

Notes: Confidence intervals are reported at the 95% level. Confidence intervals obtained via wildcluster bootstrapping. Data is from the National Incident Based Recording System, maintained by the Federal Bureau of Investigation. All estimates include a year fixed effect to capture overall trends, and month fixed effects to capture seasonal variation. Estimates are presented per month, per 100,000 residents.$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table 13: Impact of Unrestricted Concealed Carry on Non-Violent Crimes

	Armed with Firearm	Armed with Knife
Panel A: Border Counties		
Post Treatment	-0.148	-0.343
	[-0.943,0.648]	[-0.681,-0.005]
Observations	15663	15663
Mean in Trt Cnty	1.652	.973
MDE (%)	24.63	15.22
No. of Clusters	5	5
Panel B: All Counties		
Post Intervention	0.082	-0.075
	[-0.159,0.323]	[-0.291,0.142]
Observations	160511	160511
Mean in Trt Cnty	1.662	1.448
Max ES	11.45	9.66
No. of Clusters	40	40

P-Values estimated via wildcluster bootstrapping. Data is from the National Incident Based Recording System, maintained by the Federal Bureau of Investigation. Estimates are presented per month, per 100,000 residents. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table 14: Impact of Unrestricted Concealed Carry on Usage

	Murder Coef/CI	Sexual Assault Coef/CI	Aggravated Assault Coef/CI	Robbery Coef/CI
Panel A: Border Counties				
Post Treatment	-0.100	-0.112	-2.212*	0.640
	[-0.192,-0.009]	[-1.011,0.787]	[-3.603,-0.821]	[-0.053,1.334]
Observations	13892	13892	13892	13892
Mean in Trt Cnty	.274	5.299	14.069	2.146
MDE (%)	-47.81	-22.47	-49.85	-34.06
No. of Clusters	5	5	5	5
Panel B: All Counties				
Post Intervention	-0.001	-0.385	-0.539	0.070
	[-0.031,0.029]	[-1.387,0.616]	[-2.140,1.061]	[-0.155,0.295]
Observations	150094	150094	150094	150094
Mean in Trt Cnty	.239	5.4	12.494	1.476
MDE	23.06	5.95	4.39	9.69
No. of Clusters	37	37	37	37

P-Values estimated via wildcluster bootstrapping. Data is from the National Incident Based Recording System, maintained by the Federal Bureau of Investigation. Estimates are presented per month, per 100,000 residents. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 15: Restricting Analysis to LEOs that were Always Present

	Homicide Coef/CI	Sexual Assault Coef/CI	Aggravated Assault Coef/CI	Robbery Coef/CI	Burglary Coef/CI	Theft Coef/CI	Placebo Coef/CI
Post Treatment	0.062	0.035	-1.067	-0.418	1.739	0.558	-2.021
	[-0.108,0.233]	[-0.834,0.905]	[-3.656,1.521]	[-1.191,0.354]	[-5.314,8.793]	[-10.366,11.483]	[-5.617,1.575]
Observations	13892	13892	13892	13892	13892	13892	13892
Mean in Cntrl Cnty	.143	4.108	7.055	1.415	26.47	85.206	15.1
No. of Clusters	5	5	5	5	5	5	5

P-Values estimated via wildcluster bootstrapping. Data is from the National Incident Based Recording System, maintained by the Federal Bureau of Investigation. Estimates are presented per month, per 100,000 residents. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

TABLE 16: SPILLOVERS

Chapter 2. Maternal Stress, Healthcare Provision, and Child Development: Evidence from the 2015 Earthquake in Nepal

With Sajala Pandey

1 Introduction

The link between conditions during gestation and outcomes later in life has been well-documented. For example, women who were prescribed Talidomide for morning sickness in the late 1950s gave birth to babies with severe birth defects, such as missing arms and legs.[1] Jones et al. (1973) described “fetal alcohol syndrome” afflicting the children of alcoholic mothers. More recently, women are advised not to smoke or use marijuana while pregnant.[2]

Biologists have posited that prenatal maternal stress (PMNS) is a factor that can lead to adverse outcomes later in life. The theory claims that stress experienced by expectant mothers can have an adverse impact on the child , through a chain of hormonal responses. These effects can manifest both at the time of delivery and well into childhood. A strong version of this argument comes from the notion of epigenetic imprinting. The epigenome can be thought of as a series of switches that determine which parts of the genome are expressed and which parts are not. Maternal stress, in this theory, can result in the genetic imprinting of lower height and other child development variables, and the negative effects of this imprinting can last well into the adulthood. It is important to note, for the purpose of this project, that such imprinting occurs in the first trimester of the pregnancy.

Laboratory experiments provide support for this theory in animals - pregnant animals that are subject to stressors in tightly controlled laboratory conditions give birth to offspring with developmental issues. However, such studies cannot be replicated with humans subjects. Observational studies support the theory that PMNS is linked to adverse outcomes, but give rise to the problem of identification. We cannot tell if stress leads to these outcomes, or if there are some other factors, invisible to the econometrician, that lead to both stress and adverse outcomes.

One solution to this problem is to use unforeseen disasters as source of exoge-

[1] See McBride (1961) and Lenz and Knapp (1962))

[2] See Centers for Disease Control (2020).

nous variation. Disasters used by researchers to study in-utero child development include earthquakes, hurricanes, pandemics, and famines. While these papers are not always investigating stress, they typically find that these natural (and sometimes man-made) disasters have an adverse impact on children who were in-utero at the time. These findings support the theory that PMNS leads to adverse outcomes, but there are shortcomings to this methodology.

In the case of earthquakes, the event can lead to a severe disruption of the healthcare systems as hospitals are destroyed and medical practitioners injured or killed. These factors may also lead to adverse outcomes as mothers do not get adequate prenatal care, postnatal care, or a safe place to deliver their child. As a result of this, it is difficult to conclusively ascribe adverse outcomes to the biological effects of stress rather than the disruption of care induced by the disaster.

Alternatively, the event could lead to an *increase* in the amount of medical care received by mothers and children in a region, as medical aid flows into the region in response to the earthquake. Such an increase would offset the negative impact of the earthquake, causing studies to underestimate the impact of the event or even find no link.[3]

In order to tackle this problem, we study the impact that the 2015 Nepal earthquake on both child development outcomes and healthcare provision. This earthquake was unanticipated - a literal exogenous shock - and led to a great degree of death and suffering. The treatment group is the set of children[4] who were (a) *in-utero* at the time of the earthquake as well as (b) in areas that were severely affected by the earthquake. The control group consists of children who do not fulfill one or more of the aforementioned conditions. We further divide the treatment group into children whose mothers received regular antenatal care and those whose mothers did not.

Our household and biomarker data are drawn from the 2016 Nepal Demographic and Health Survey (DHS), a nationally representative survey that covers over 11,000 households. Enumerators, with parental consent, measure and record height and weight of children in the household. The DHS reports the height-for-age of these children, and this variable is our primary measure of interest. These data also include the number of antenatal visits taken by the mother during the pregnancy, which we use as a proxy for the amount of antenatal medical care received. The survey also includes a variety of household characteristics which we use as controls in our analysis.

We find that children who were in-utero at the time of the earthquake were shorter on average than their cohort. The distribution of these effects are consistent with PNMS theories. We further divide the control group into those who were in the

[3]Jamous (2020), for example, finds no link between a series of earthquakes in Chile and child development outcomes.

[4]"Children" here refers to people below the age of 5 in our data.

first trimester when the earthquake struck - this stage being critical for fetal development and imprinting - and those who were not. We find that the negative effects of the earthquake were concentrated among those who were in the first trimester at the time, and those in the second or third were not affected, consistent with the theory from imprinting and PNMS.

A potential concern in any analysis of a natural disaster is "culling" - a selection effect in which the weakest members of a population are the most likely to be killed. In this data, that could result in an increase in infant mortality amongst the weakest children or miscarriages among the weakest mothers. Such an effect, in the absence of any other, would cause our analysis to report an *improvement* in the developmental outcomes of children in our data. Our analysis would understate the deleterious impact of the earthquake. The above effect should therefore be treated as a floor, rather than an accurate point estimate.

We find that receiving antenatal care offsets the negative effects of the earthquake. Interestingly, the magnitudes and distribution of this effect was such that they canceled out the negative effects caused by the earthquake in our data, and we found no overall impact of the earthquake on the treatment group as a whole. This conclusion should be interpreted with care, as it is possible that the getting healthcare was correlated with some unobserved household or individual characteristic.

In conclusion, we find support for the theory that PNMS in the first trimester can be detrimental to child development. We also find, with some caveats, evidence that the proper provision of antenatal healthcare can offset these detrimental effects, and that the provision of medical care can be extremely important to early childhood development, even after the child is born.

2 Background & Literature

2.1 The earthquake in Nepal

Nepal is a landlocked country of 28 million people located in South Asia. It borders Tibet in the north and India in the south, east, and west. Its geographic boundaries include a substantial portion of the Himalaya mountain range, which is formed by the subduction of the Indian tectonic plate under the Eurasian Plate. This subduction makes the country potentially vulnerable to devastating earthquakes.[5] Prior to the 2015 earthquake, the worst natural disaster in Nepal's history was the 1934 Bihar-Nepal earthquake.

The particular earthquake this paper studies occurred on the 25th of April, 2015 in the Gorkha District of Nepal, and had a magnitude of 7.8 on the moment

[5] See Bilham et al. (2004)

magnitude scale.[6] In total, it killed nearly 9000 people, injured 22,000, and left 3.1 million people homeless. The death toll would have been higher had the earthquake not happened at noon on a Saturday, when schools were closed and a substantial portion of the country were working in open fields. It was followed by a major aftershock 17 days later on 12 May 2015 at 12:50pm, which killed over 200 people and injured 2500.

Following the earthquake, the Nepalese army was dispatched to affected areas, and volunteers were arrived from other parts of the country. International aid to Nepal totaled $3 billion USD in cash. This was in addition to material aid (food, water etc.) and rescue aid (medical personnel, aircraft etc.)

2.2 Maternal Stress & Child Development

Beydoun and Saftlas (2008) review the literature on the link between PNMS and developmental outcomes, and find broad support for the theory that the two are inversely linked. There are two types of studies in this field: animal studies and human studies. In animal studies, pregnant animals are subject to stressful stimuli and the impact on their offspring is noted. The animals were usually rats or primates and they were subject to stressors such as restraints, acoustic startles, bright lights, and electric shocks, among others. These stressors tended to have adverse effects on their offspring's development, including their growth, sexual maturation, motor development, and immune responses.

Such experiments cannot be replicated with human subjects, for reasons that are self-evident. Therefore, for human studies, researchers typically adopt an observational approach, recording the stress levels of expectant mothers[7] and relating it to outcomes such as length of gestation, risk of preterm labor, or preterm birth. Such studies cannot demonstrate causality - it isn't clear whether the stress caused these adverse outcomes, or some third factor led to both stress and these outcomes.

A natural question to ask is: through what mechanism does the stress lead to bad outcomes? One possibility is that PNMS induces preterm delivery though hormones released by the hypothalamic–pituitary–adrenal (HPA) axis[8]. This may lead to adverse birth outcomes, such as lower birth weight, that may affect outcomes in the future. This mechanism is of particular note because several primate studies involve the injection of hormones directly into the subjects. In the later stages of pregnancy, the placenta prevents these hormones from affecting the child, but the protection is weakest during the first trimester.

[6] A technical account of the earthquake can be found in Goda et al. (2015)

[7] Measures include stressful life events, perceived stress, anxiety, and prenatal distress.

[8] In response to a stressful situation, the hypothalamus secretes corticotropin-releasing hormone (CRH) which stimulates the pituitary gland, which produces adrenocorticotropin hormone (ACTH). This in turn stimulates the adrenal cortex which produces cortisol. In primates, the presence of cortisol stimulates the placenta which produces placental CRH (pCRH) via a positive feedback loop. This hormone plays an role in regulating parturition late in the pregnancy, and abnormally high pCRH is associated with increased risk of preterm delivery

A more controversial answer draws on the "fetal origins hypothesis". This hypothesis - which posits that events in early life can be linked to outcomes in later life - is also known as the Barker hypothesis after one of it's leading proponents. In an influential paper, Barker and Osmond (1986) argued that that poor nutrition in-utero can lead to obesity and cardiovascular disease later in life, using infant mortality and cardiovascular mortality data from counties in England and Wales. A shortcoming of Barker's work is that it relies heavily on correlational studies and did not adequately adjust for confounders, leading to much discussion on the validity of their findings. Almond and Currie (2011) provide an excellent overview of their work and the issues surrounding it.

Much of the discourse centers around the notion that prenatal exposure to certain environments can lead to "programming" or "imprinting" of the epigenome. The epigenome can be thought of as a series of switches that determine which parts of the genome are expressed and which parts are not (Petronis 2010). Maternal stress, in this argument, can result in the genetic imprinting of lower height and other child development variables. It is important to note, for the purpose of this project, that such imprinting occurs in the first trimester of the pregnancy.

Both of these hypothetical channels have something in common: they are at their most influential during the first trimester. Thus, children whose mothers were subject to stress during the first trimester should have worse developmental indicators than children whose mothers were subject to stress in the second or third trimester.

The impact of prenatal care on maternal health is an under-studied topic. Studies find little to no effect in developed countries,[9] but significant effects in developing countries.[10]

2.3 Natural disasters as natural experiments

As discussed in the previous section, it is difficult for observational studies to conclusively demonstrate that stress causes adverse outcomes. As a result, researchers sometimes use unforeseen disasters as source of exogenous variation in stress. Disasters used by researchers to study in-utero child development include earthquakes,[11] hurricanes,[12] pandemics,[13] and famines.[14] While these papers do not all investigate

[9] See, for example, Noonan et al. (2013)

[10] See, for example, X. Liu et al. (2017)

[11] See Torche (2011), Palmeiro-Silva et al. (2018), Berthelon, Kruger, and Sanchez (2018), Menclova and Stillman (2019), Jamous (2020), Caruso and Miller (2015), Guo et al. (2019), Glynn et al. (2001), and Kim, Carruthers, and Harris (2017). Of particular interest is Paudel and Ryu (2018) who examine the impact of the 1988 earthquake in Nepal and find that it led to a decrease in human capital, especially among children of lower caste groups.

[12] See Karbownik and Wray (2019) and Sotomayor (2013)

[13] The seminal paper in this area is arguably Almond (2006), who uses the 1918 influenza epidemic in the US to investigate the long term effects of in-uterus exposure, and finds that it led to fewer years of schooling, a higher incidence of disability, and lower wages. Also see Lin and E. M. Liu (2014) and Barreca (2010)

[14] Almond, Edlund, et al. (2010) look at the 1959-61 Chinese famine and find increased rates of disability and non-participation in the labor market. Scholte, Van Den Berg, and Lindeboom (2015) look at the Dutch hunger

stress specifically, they typically find that these natural (and sometimes man-made) disasters have an adverse impact on children who were in-utero at the time. These findings support the theory that PNMS leads to adverse outcomes, but there are shortcomings to such an approach.

Earthquakes, for example, can lead to a severe disruption of the healthcare systems as hospitals are destroyed and medical practitioners injured or killed. As a result of this, mothers may not have access to adequate antenatal care or a safe place to deliver their child, and children may not receive the early life medical care they need for their development. Alternatively, earthquakes may lead an *increase* in the amount of medical care that some households receive, as medical assistance floods into the area along with other forms of aid. Thus studies may fail to find an effect even when it exists, or understate the effect that exists, if they do not account for this possibility.

We address this problem by not only analyzing the impact of the 2015 Nepal earthquake on child outcomes, but also on the quality and quantity of care received by mothers in areas that were severely hit by the earthquake. We are enabled in this by access to high quality data about children's outcome and the antenatal care received by their mothers.

3 Data

Household and biomarker data are drawn from the 2016 Nepal Demographic and Health Survey (NDHS) conducted by the United States Agency for International Development (USAID) in collaboration with the Ministry of Health of the Government of Nepal. The NDHS is a nationally-representative household survey that is conducted every 5 years to monitor the changes in population, health, and nutrition. The 2016 survey was the fifth of its' kind to be conducted in Nepal.

The 2016 survey covers 11,040 households in Nepal. The target groups were women and men age 15-49 residing in randomly selected households across the country, and the sample is designed to represent the population in this age group. The survey consists of a household questionnaire (administered to a woman in the household), a women's questionnaire, and a men's questionnaire. Enumerators - with the consent of the parents of the child - check the weight and height of children below the age of 5. This section - known as a the *children's recode* - is the primary focus of this study. The children's recode also includes information about the child's sex, their birth order, and the number of siblings.

Additionally, enumerators asked women about the prenatal and postnatal medical care they received, such as the number of prenatal medical checkups, the num-

winter in the Netherlands and find a significant impact on employment. Jürges (2013) and Neelsen and Stratmann (2011) also look at European famines. On a related note,Almond and Mazumder (2011) look at the impact of fasting during Ramadan.

ber of postnatal baby checks, immunizations, the place of delivery, and so on. Enumerators asked mothers whether their child's size at birth was, in their subjective opinion, higher or lower than normal.

The survey includes a roster of residents, which we used to generate the household size variable. It also includes the education level of each member of the household, the highest of which is included as proxy measure of the level of education in the household. The roster includes information about the children being analyzed, including their gender, birth order, and number of siblings.

Women were asked about their age and level of education, as well as their partner's age and level of education. These variables were used to control for mother's and father's characteristics in our analysis.

The survey elicited information about the ownership of various assets in the household, which was used - by the NDHS - to construct a wealth index. We used this index as a measure of household wealth. An indicator also records whether the household was in a rural area or not.

The survey recorded the biomarkers of 2,379 children, from a sample of 5,038, and these observations are the primary focus of this paper. Table 1 presents the summary statistics for these children.

Data on the severity of the earthquake are drawn from the US Geological Survey. Affected areas were mapped to this data by mapping village-and-district codes to their GPS co-ordinates.

4 Methodology

In this section, we outline the methodology used to analyze our data.

4.1 Impact of the earthquake and healthcare provision

We estimate the following equation:

$$Y_{ivd} = \alpha + \beta_1 D_i + \beta_2 s_v + \beta_3 c_{ivd} + \beta_4 (D_i \times s_v) + \beta_5 (D_i \times s_v \times c_{ivd}) + \delta X_{iv} + \gamma_d + \epsilon_{ivd} \quad (1)$$

where Y_{ivd} is the outcome of interest for child i in village v and district d, D_i is a binary variable that takes the value 1 if the child was in-utero at the time of the earthquake and 0 otherwise, s_v is a variable that takes the value 1 if the earthquake was classified as being severe in their region (by the US Geological Survey) and 0 otherwise, c_{ivd} which takes the value 1 if the mother went to 4 or more antenatal visits[15], γ_d is a set of district-level fixed effects, and X_{iv} is a set of controls.

These controls include: the child's gender, birth order, and number of siblings; mother's age and education; father's age and education; the size of the household,

[15]This number was chosen because 4 is the median number of antenatal visits in our dataset.

a household wealth index, whether the household was located in a rural or urban area, and the highest level of education in the household.

We are interested in the direction and magnitude of two coefficients: β_4 and β_5. β_4 measures the impact of the earthquake on children who were *in-utero* at the time, whose mothers were physically in severely affected areas, and whose mothers received fewer than 4 antenatal visits. β_5 measures the impact of the earthquake on children who were *in-utero* at the time, whose mothers were physically in areas considered to be severely affected by the earthquake, and whose mothers received 4 or more antenatal visits. The latter allows us to measure the heterogenous impact of the earthquake on those who did not or could not get antenatal care.

If the term $D_i \times s_v \times c_{ivd}$ were to be omitted from the specification, the estimated β_4 would actually be reporting the weighted average of β_4 and β_5. If the two coefficients had different signs then the specification would underestimate β_4.

Standard errors are clustered at the district level.

4.2 Breakdown by trimester

As we discussed earlier, the epigenetic and hormonal theories that link PMNS to adverse development outcomes predict that the first trimester is the most critical to the child's development, therefore we are interested in whether children who were in the first trimester at the time of the earthquake show worse outcomes than those who were in later stages.

To test this, we estimate the following equation:

$$Y_{ivd} = \alpha + \beta_{1j} T_{ij} + \beta_2 s_v + \beta_{3j}(T_{ij} \times s_v) + \delta X_{iv} + \gamma_d + \epsilon_{ijvd} \tag{2}$$

where Y_{iv} is the outcome variable of interest for child i in village v and district d, T_{ij} are a series of binary variables that take the value 1 if the child was in-utero in trimester $j \in \{1, 2, 3\}$ at the time of the earthquake and 0 otherwise, s_v is a variable that takes the value 1 if the earthquake was classified as severe in their region and 0 otherwise, γ_d is a set of district-level fixed effects, and X_{iv} is a set of household level controls and individual controls. In this specification, the three β_{3j} are the coefficients of interest, showing the impact of the earthquake on outcomes if the child was in trimester j at the time of the earthquake. The controls are the same as in Equation 1, and standard errors are clustered at the district level.

5 Results

Our primary outcome of interest is a variant of height for age provided by the DHS: the number of standard deviations from a reference median. We normalize this variable before deploying it in our analysis.

The first question we tackle is whether the earthquake was, in fact, detrimental to developmental outcomes. Column (3) of Table 2 presents the estimated coefficients of Equation 1 for normalized height for age. The coefficient corresponding to the interaction of the severity and in-utero dummy variables is negative and statistically significant, supporting the theory that the earthquake was detrimental to child development. Due to the possibility of the weakest children being "culled" by the earthquake and its aftermath, the point estimate should be treated as a floor, rather than an accurate point estimate, of the impact of this earthquake.

In the same table, the coefficient corresponding to the interaction of severity, in-utero, and regular ante-natal care dummy variables is positive and statistically significant. This lends support to the theory that ante-natal care is beneficial to child development. Note that the two coefficients are nearly identical, suggesting that sufficient antenatal care could not only ameliorate, but even nullify the negative effects of the earthquake. Support for this interpretation can be seen in column (2). In this specification, we drop the dummy variable that corresponds to the interaction of severity, in-utero, and regular antenatal care dummy variables. As a result the coefficient on the interaction of the severity and in-utero dummies is now statistically insignificant. This variable now includes the effect of the earthquake on both those whose mothers received regular antenatal care, and those who didn't, and the overall effect is no longer detectable.

This interpretation should, however, be considered with care. An alternate possibility that getting regular antenatal care is related with some unobserved variable (through socioeconomic status or accessibility) that is responsible for ameliorating effect.

How did the earthquake affect the provision of healthcare overall? In the short term, the earthquake almost surely reduced this supply, as hospitals collapsed and medical personnel died. In the medium term, the impact is less clear, as aid effort floods these areas with volunteers, many of them medical personnel.

The direction of the overall effect is hinted by a striking fact: children who were in severely-affected areas, but not in-utero at the time, were *taller* than their cohort, as evidenced by the positive coefficient on the severe coefficient. (These include both children who were born before the earthquake or conceived after it.) This result suggests that healthcare provision *rose* overall, to the benefit of young children.

To test this possibility, we can look directly at the impact of the earthquake on healthcare provision. Table 4 presents the estimates of Equation 1 for outcomes related to the provision of healthcare. We see that the earthquake had no impact on at-home births, and that children in these areas were *more* likely to be immunized and more likely to enjoy postnatal checks than their cohort in other areas; this is true even for children who were in utero at the time. As evidenced from column (1) children who were born in this area but conceived after the earthquake, enjoyed more antenatal visits than their cohort.

We should nonetheless be careful in reaching this conclusion. These observations are also consistent with with "culling" - the possibility that the earthquake disproportionately kills weaker children or parents, with the consequence that those left are taller on average.

(The earthquake also seems to have hurt children who were not in severely affected areas, possibly by causing economic turmoil in Nepal. Evidence for this can be seen Column (3) of Table 2 where the coefficient corresponding to the in-utero dummy variable is negative. While interesting, this line of inquiry is beyond the scope of this paper.)

None of the findings that we've so far discussed support or negate the theory that PNMS causes adverse child developmental outcomes. To test this theory, we look at the impact the earthquake had on children in different trimesters. The estimates for Equation 2 is presented in Column (3) of Table 3. The coefficient on the interaction variable is negative and statistically significant for those in the first trimester, but not the second or third. This is consistent with the theory that stressors in the first trimester of pregnancy have an acute effect, possibly due to epigenetic imprinting.

It is important to note that the coefficient on the triple interaction variable - seen in Column (4) - is positive and statistically significant, suggesting that antenatal care can ameliorate the impact of the stressors and disasters. Consistent with this interpretation, the coefficient of the interaction variable is now larger - almost double - in magnitude. Once again, these conclusions must be reached carefully, as the reception of antenatal care could be the result of some unobserved variable.

We break the provision of antenatal visits down by trimester in Table 5. Here we observe that the while the earthquake reduce the number of antenatal visits for those who were late in their pregnancy, it *increased* it for those who were in the first trimester when it struck. In other words, the negative effects of the earthquake on those who were in the first trimester occurred *despite* the healthcare-provision related effects of this earthquake, lending credence to PMNS theory of child development.

6 Robustness checks

We look at alternative ways to specify height. Tables 6 and 7 present the estimates of Equation 1 for height for age as a percentile and as a percent of reference median instead. Each of these variables have been normalized such that they have an average of 0 and an standard deviation of 1. The direction of the coefficients is identical to that found in the previous section, and the coefficients are similar in their statistical significance.

Next, we consider alternate measures. Table 8 presents the estimates for Equation 1 for the outcome variable normalized weight for age. The estimates are in the same direction to our estimates for height for age, and are similarly statistically significant. The interpretation remains the same - children who were in severe areas but not in utero were *heavier* than others in their cohort, being in utero at the time of the earthquake was detrimental to development, and the reception of antenatal care helped lower the detrimental effects of being in-utero at the time.

7 Discussion

One significant takeaway from this analysis is that healthcare provision in the aftermath of a natural disaster is critical. The provision of such care, in our data, not only ameliorates but apparently offsets the negative impact of the earthquake on those who received it. With caveats about culling, the provision of such care seems to have *improved* the outcome of children who weren't in-utero at the time and were therefore beneficiaries of the influx of medical aid from other parts of Nepal and the rest of the world.

This conclusion, by itself, neither supports nor disproves theories about the potential impact of PNMS on child outcomes. Such theories predict that the impact of stress should be most severe amongst children whose mothers were in the first trimester of their pregnancy. We find that the impact of the earthquake was most severe among children whose mothers were in the first trimester of their pregnancy. Our analysis supports theories about the potential impact of PNMS, with the caveat that we cannot rule out some other factors that may be at work here.

We have previously discussed the hormone channel and epigenetic channel through which PNMS can affect outcomes. Beydoun and Saftlas (2008) note two additional possible biological channels that link stress and child development. First, stress during pregnancy can increase vulnerability to maternal infections via mechanisms that inhibit components of the immune system. Second, stress can lead to unhealthy behavior such as substance abuse, or poor nutrition, which may have adverse effects on the child. We find no evidence to support these theorized channels which, to the best of our knowledge, do not affect women in the first trimester differently than women in the second or third.

8 Conclusion

Economic outcomes in later life are strongly affected by shocks in the early stages of development, including pregnancy. Biologists have theorized that a link exists between prenatal maternal stress and child developmental outcomes, but such theories are difficult to test in a laboratory setting with human subjects.

Scientists have attempted to test the impact of shocks on child development outcomes using natural disasters as an exogenous source of variation. A shortcoming in these approaches - for the investigation of PNMS - is that shocks can also lead to the changes in the supply of healthcare. In the very short run, such shocks will lead to a shortage of healthcare, as hospitals are destroyed overwhelmed, and medical practitioners are killed or injured. In the medium run, shocks can lead to an increase in the supply of healthcare, as medical aid flows into the region in response to the disaster. Such variation may lead to erroneous conclusions about the impact of shocks.

We analyze the impact of the 2015 Earthquake in Nepal on the height of children who were in severely affected areas and in-utero at the time. Expanding the previous literature, we control for the availability of healthcare using antenatal visits as a proxy. We find that the earthquake had a negative impact on the developmental outcomes of the control group, that these outcomes were apparently ameliorated by the provision of healthcare, and that these two effects cancelled each other out in our preferred specification. The ameliorating effects of healthcare should be carefully considered, as they are consistent with the possibility the existence some unobserved factor that is correlated with their consumption.

Consistent with theories that link PNMS to child outcomes, we found that the negative effects of the earthquake were concentrated on the children who were in the first trimester at the time of development.

Overall, our analysis supports the theory that there exists a biological link between PNMS and adverse child outcomes, and highlights the potential importance of healthcare policy in breaking it.

References

Almond, Douglas (2006). "Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population". In: *Journal of political Economy* 114.4, pp. 672–712.

Almond, Douglas and Janet Currie (2011). "Killing me softly: The fetal origins hypothesis". In: *Journal of economic perspectives* 25.3, pp. 153–72.

Almond, Douglas, Lena Edlund, et al. (2010). "Long-term effects of early-life development: Evidence from the 1959 to 1961 china famine". In: *The economic consequences of demographic change in East Asia*. University of Chicago Press, pp. 321–345.

Almond, Douglas and Bhashkar Mazumder (2011). "Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy". In: *American Economic Journal: Applied Economics* 3.4, pp. 56–85.

Barker, David JP and Clive Osmond (1986). "Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales". In: *The Lancet* 327.8489, pp. 1077–1081.

Barreca, Alan I (2010). "The long-term economic impact of in utero and postnatal exposure to malaria". In: *Journal of Human resources* 45.4, pp. 865–892.

Berthelon, Matias, Diana I Kruger, and Rafael Sanchez (2018). "Maternal stress during pregnancy and early childhood development". In:

Beydoun, Hind and Audrey F Saftlas (2008). "Physical and mental health outcomes of prenatal maternal stress in human and animal studies: a review of recent evidence". In: *Paediatric and perinatal epidemiology* 22.5, pp. 438–466.

Bilham, Roger et al. (2004). "Earthquakes in India and the Himalaya: tectonics, geodesy and history". In: *Annals of GEOPHYSICS*.

Caruso, Germán and Sebastian Miller (2015). "Long run effects and intergenerational transmission of natural disasters: A case study on the 1970 Ancash Earthquake". In: *Journal of development economics* 117, pp. 134–150.

Centers for Disease Control (2020). "Pregnant? Don't Smoke!" In: *Monograph*.

Glynn, Laura M et al. (2001). "When stress happens matters: effects of earthquake timing on stress responsivity in pregnancy". In: *American journal of obstetrics and gynecology* 184.4, pp. 637–642.

Goda, Katsuichiro et al. (2015). "The 2015 Gorkha Nepal earthquake: insights from earthquake damage survey". In: *Frontiers in Built Environment* 1, p. 8.

Guo, Chao et al. (2019). "Long-term effects of prenatal exposure to earthquake on adult schizophrenia". In: *The British Journal of Psychiatry* 215.6, pp. 730–735.

Jamous, Christoffer Karlsson (2020). *Shaken by the stress: Does in-uterus earthquake exposure cause long-term disadvantages for the fetus?*

Jones, KennethL et al. (1973). “Pattern of malformation in offspring of chronic alcoholic mothers”. In: *The Lancet* 301.7815, pp. 1267–1271.

Jürges, Hendrik (2013). “Collateral damage: The German food crisis, educational attainment and labor market outcomes of German post-war cohorts”. In: *Journal of Health Economics* 32.1, pp. 286–303.

Karbownik, Krzysztof and Anthony Wray (2019). “Long-run consequences of exposure to natural disasters”. In: *Journal of Labor Economics* 37.3, pp. 949–1007.

Kim, Bongkyun, Celeste K Carruthers, and Matthew C Harris (2017). “Maternal stress and birth outcomes: Evidence from the 1994 Northridge earthquake”. In: *Journal of Economic Behavior & Organization* 140, pp. 354–373.

Lenz, Widukind and Klais Knapp (1962). “Thalidomide embryopathy”. In: *Archives of Environmental Health: An International Journal* 5.2, pp. 14–19.

Lin, Ming-Jen and Elaine M Liu (2014). “Does in utero exposure to illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment”. In: *Journal of health economics* 37, pp. 152–163.

Liu, Xiaoying et al. (2017). “Prenatal care and child growth and schooling in four low-and medium-income countries”. In: *PloS one* 12.2, e0171299.

McBride, William Griffith (1961). “Thalidomide and congenital abnormalities”. In: *Lancet* 2.1358, pp. 90927–8.

Menclova, Andrea Kutinova and Steven Stillman (2019). *Maternal Stress and Birth Outcomes: Evidence from an Unexpected Earthquake Swarm*. Tech. rep. IZA Discussion Papers.

Neelsen, Sven and Thomas Stratmann (2011). “Effects of prenatal and early life malnutrition: Evidence from the Greek famine”. In: *Journal of Health Economics* 30.3, pp. 479–488.

Noonan, Kelly et al. (2013). “Effects of prenatal care on child health at age 5”. In: *Maternal and child health journal* 17.2, pp. 189–199.

Palmeiro-Silva, Yasna K et al. (2018). “Effects of earthquake on perinatal outcomes: A Chilean register-based study”. In: *PloS one* 13.2, e0191340.

Paudel, Jayash and Hanbyul Ryu (2018). “Natural disasters and human capital: The case of Nepal?s earthquake”. In: *World Development* 111, pp. 1–12.

Petronis, Arturas (2010). “Epigenetics as a unifying principle in the aetiology of complex traits and diseases”. In: *Nature* 465.7299, pp. 721–727.

Scholte, Robert S, Gerard J Van Den Berg, and Maarten Lindeboom (2015). “Long-run effects of gestation during the Dutch Hunger Winter famine on labor market and hospitalization outcomes”. In: *Journal of health economics* 39, pp. 17–30.

Sotomayor, Orlando (2013). “Fetal and infant origins of diabetes and ill health: Evidence from Puerto Rico’s 1928 and 1932 hurricanes”. In: *Economics & Human Biology* 11.3, pp. 281–293.

Torche, Florencia (2011). “The effect of maternal stress on birth outcomes: exploiting a natural experiment”. In: *Demography* 48.4, pp. 1473–1491.

	Mean	Std Dev
Outcomes		
Height for Age (standard deviations from the reference median)	-0.39	10.00
Weight for Age (standard deviations from the reference median)	-0.49	9.98
Size of the Child	3.05	0.80
Child Characteristics		
Female Child	0.48	0.50
Birth Order	2.26	1.54
Number of Siblings	1.45	1.58
Parental Chracteristics		
Mother's Age	26.24	5.62
Mother's Education: Secondary or Higher	0.48	0.50
Father's Education: Secondary or Higher	0.63	0.48
Father's Age	30.33	6.76
Household Characteristics		
Household Size	6.12	2.88
Wealth Index	2.72	1.35
Rural Location	0.44	0.50
Highest Education	7.86	6.51
Care		
Antenatal Visits	4.20	2.23
Postnatal Baby Check	0.37	0.54
No of Immunizations	0.74	0.44
Delivery at Home	0.41	0.49
Observations	2379	

Notes: Data drawn from the Nepal Demographic and Health 2016.

TABLE 1: SUMMARY STATISTICS FOR SAMPLE

	(1)	(2)	(3)
Severe	0.26***	0.26***	0.27***
	(0.06)	(0.06)	(0.06)
Severe x In Utero	-0.13	-0.14	-0.37**
	(0.08)	(0.09)	(0.16)
In Utero	-0.14***	-0.14***	-0.14***
	(0.04)	(0.04)	(0.04)
Regular Care		-0.02	-0.02
		(0.04)	(0.04)
Severe x In Utero x Regular Care			0.39**
			(0.14)
District Fixed Effects	Yes	Yes	Yes
Household Characteristics	Yes	Yes	Yes
Individual Effects	Yes	Yes	Yes
Regular ANC	No	Yes	Yes

Notes: Data drawn from the Nepal Demographic and Health 2016. Household characteristics include size, wealth index, and the highest level of educaiton. Individual characteristsics include gender, mother's education, and mother's age. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 2: Height for Age (Normalized Standard Deviations from the Reference Median)

	(1)	(2)	(3)	(4)
Severe	0.26***	0.25***	0.27***	0.27***
	(0.06)	(0.05)	(0.06)	(0.06)
Severe x In Utero	-0.13			
	(0.08)			
In Utero	-0.14***			
	(0.04)			
In First Trimester		-0.17***	-0.16***	-0.16***
		(0.05)	(0.05)	(0.05)
In Second Trimester		-0.14**	-0.14**	-0.14**
		(0.06)	(0.06)	(0.06)
In Third Trimester		-0.14**	-0.14**	-0.14**
		(0.05)	(0.05)	(0.05)
In First Trimester x Severe			-0.21**	-0.40***
			(0.08)	(0.12)
In Second Trimester x Severe			-0.06	-0.07
			(0.08)	(0.08)
In Third Trimester x Severe			-0.03	-0.06
			(0.18)	(0.18)
Regular Care				-0.02
				(0.04)
Severe x In First Trimester x Regular Care				0.42***
				(0.14)
District Fixed Effects	Yes	Yes	Yes	Yes
Household Characteristics	Yes	Yes	Yes	Yes
Individual Effects	Yes	Yes	Yes	Yes
Care Proxies	No	Yes	Yes	Yes

Notes: Data drawn from the Nepal Demographic and Health 2016. Household characteristics include size, wealth index, and the highest level of educaiton. Individual characteristsics include gender, mother's education, and mother's age. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 3: Height for Age by Trimester

	Antenatal Visits	At-home Birth	Immunized	Postnatal Check
Severe	0.53***	0.07	0.21***	0.28***
	(0.10)	(0.05)	(0.03)	(0.05)
Severe x In Utero	-0.45***	0.04	0.53***	0.43***
	(0.11)	(0.10)	(0.12)	(0.12)
In Utero	0.07	-0.05	0.71***	-0.06
	(0.05)	(0.03)	(0.05)	(0.06)
District Fixed Effects	Yes	Yes	Yes	Yes
Household Characteristics	Yes	Yes	Yes	Yes
Individual Effects	Yes	Yes	Yes	Yes

Notes: Data drawn from the Nepal Demographic and Health 2016. Household characteristics include size, wealth index, and the highest level of educaiton. Individual characteristsics include gender, mother's education, and mother's age. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

TABLE 4: CARE RECIEVED

	(1)	(2)	(3)	(4)
Severe	1.10***	0.53***	1.10***	0.51***
	(0.09)	(0.10)	(0.09)	(0.09)
Severe x In Utero	-0.30*	-0.45***		
	(0.15)	(0.11)		
In Utero	0.07	0.07		
	(0.06)	(0.05)		
In First Trimester x Severe			0.95***	0.94***
			(0.15)	(0.14)
In Second Trimester x Severe			-0.99***	-1.41***
			(0.10)	(0.14)
In Third Trimester x Severe			-1.86***	-2.07***
			(0.21)	(0.21)
In First Trimester			-0.11	-0.07
			(0.13)	(0.10)
In Second Trimester			0.08	0.12
			(0.08)	(0.11)
In Third Trimester			0.25	0.15
			(0.15)	(0.11)
District Fixed Effects	No	Yes	No	Yes
Household Characteristics	No	Yes	No	Yes
Individual Effects	No	Yes	No	Yes
Care Proxies	No	No	No	No

Notes: Data drawn from the Nepal Demographic and Health 2016. Household characteristics include size, wealth index, and the highest level of educaiton. Individual characteristsics include gender, mother's education, and mother's age. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 5: Antenatal Visits

	(1)	(2)	(3)
Severe	0.81^{***}	0.81^{***}	0.82^{***}
	(0.04)	(0.04)	(0.04)
Severe x In Utero	-0.76^{***}	-0.73^{***}	-1.06^{***}
	(0.06)	(0.08)	(0.16)
In Utero	-0.32^{***}	-0.33^{***}	-0.33^{***}
	(0.06)	(0.06)	(0.06)
Regular Care		0.10	0.10
		(0.07)	(0.07)
Severe x In Utero x Regular Care			0.54^{***}
			(0.16)
District Fixed Effects	Yes	Yes	Yes
Household Characteristics	Yes	Yes	Yes
Individual Effects	Yes	Yes	Yes
Regular ANC	No	Yes	Yes

Notes: Data drawn from the Nepal Demographic and Health 2016. Household characteristics include size, wealth index, and the highest level of educaiton. Individual characteristsics include gender, mother's education, and mother's age. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$

Table 6: Height for Age (Percentile)

	(1)	(2)	(3)
Severe	0.23***	0.23***	0.23***
	(0.06)	(0.06)	(0.06)
Severe x In Utero	-0.10	-0.12	-0.34*
	(0.08)	(0.09)	(0.16)
In Utero	-0.12**	-0.12**	-0.12**
	(0.04)	(0.04)	(0.04)
Regular Care		-0.03	-0.03
		(0.04)	(0.04)
Severe x In Utero x Regular Care			0.37**
			(0.13)
District Fixed Effects	Yes	Yes	Yes
Household Characteristics	Yes	Yes	Yes
Individual Effects	Yes	Yes	Yes
Regular ANC	No	Yes	Yes

Notes: Data drawn from the Nepal Demographic and Health 2016. Household characteristics include size, wealth index, and the highest level of educaiton. Individual characteristsics include gender, mother's education, and mother's age. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$

Table 7: Height for Age (Percent of Reference Median)

	(1)	(2)	(3)
Severe	0.24***	0.24***	0.25***
	(0.06)	(0.06)	(0.06)
Severe x In Utero	-0.11	-0.11	-0.44**
	(0.09)	(0.09)	(0.19)
In Utero	-0.15***	-0.15***	-0.15***
	(0.04)	(0.04)	(0.04)
Regular Care		-0.02	-0.03
		(0.04)	(0.04)
Severe x In Utero x Regular Care			0.53***
			(0.16)
District Fixed Effects	Yes	Yes	Yes
Household Characteristics	Yes	Yes	Yes
Individual Effects	Yes	Yes	Yes
Regular ANC	No	Yes	Yes

Notes: Data drawn from the Nepal Demographic and Health 2016. Household characteristics include size, wealth index, and the highest level of educaiton. Individual characteristsics include gender, mother's education, and mother's age. $^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$

Table 8: Weight for Age (Normalized Standard Deviations from the Reference Median)

Chapter 3. Multiunit Allocation in Small Economies

With Andrew Copland

1 Introduction

Registering for college courses can be stressful, as many universities use a first-come-first-served (FCFS) approach to assigning students to classes. Students who register early get their pick of classes while those who register late don't, sometimes for reasons that are no fault of their own. Furthermore, students adopt wasteful strategies to ensure they get their choice of classes, including staying up late, waking up early, or even writing bots to handle registration for them[1].

An alternative to FCFS is a lottery that gives winners priority registration, also known as *randomized serial dictatorship*, a method often used by many US universities to assign dorm rooms to students. While this reduces the deadweight loses associated with FCFS, this too runs into the above problem: winners get all of their preferred classes, while losers get none. It is unfair *ex-post*.

This has prompted some schools to use a *simple draft*. Students submit a ranking of their courses, and a lottery for ranks is held. The top ranked student is assigned their top choice, the second gets their top choice from the options left, and so on. At the end of the first round, ranks are reversed and this process is repeated until all classes or all schedules are full.

This process is manipulable - students know that they should over-rank popular courses and under-rank unpopular courses to maximize their chances of getting into in-demand courses. Budish and Cantillon (2012) analyze a draft conducted by Harvard Business School, comparing stated and true preferences, and show that not only does this manipulation happen, but that the strategizing leads to severe welfare loss. Popular courses reach maximum capacity quickly, hurting those students who sincerely preferred them, and students may end up not getting the courses they actually wanted because they strategically under-ranked them.[2]

Another popular solution, especially amongst business schools, is to hold an auction for classes using some form of invented currency. While such auctions are fair and efficient in theory, in practice they are often confusing, time consuming, opaque, and manipulable. Prices can vary tremendously from year to year for rea-

[1] See McKenzie (2019) for a discussion of this phenomena

[2] These outcomes were nonetheless better than randomized serial dictatorship.

sons that are unclear, and business schools often issue lengthy guides to help students navigate such markets.[3]

We propose an algorithm that we will refer to as a *revision draft*. Students submit their preferences as an ordering, a lottery for ranks is held, and a simple draft is conducted. Then the algorithm will analyze each student in the order of their rank. For each student, the algorithm will consider the universe of orderings they could have submitted instead, and what their outcomes would have been if they had. If the student would do better by submitting one of those orderings, the algorithm replaces their current ordering with the alternative. This process is repeated until no student can further improve their current allocation with an ordering.

Our algorithm has properties desirable in a matching mechanism. We prove that, when it converges without encountering a loop, it is strategy-proof[4] which is valuable in contexts where strategizing imposes deadweight costs and where some agents are less savvy than others. We prove that it is efficient[5] i.e. that there are no unrealized gains that could have been enjoyed. It is non-dictatorial, retains all properties in small markets, and is relatively easy to understand and implement.

A drawback of this mechanism is that, in any reasonable-sized market, the level of computing power required by this algorithm is astronomical - the amount of time required to analyze all options for a single student would exceed the expected lifetime of the universe. In order to tackle this problem we propose that, in simulations, students be restricted to manipulating the top n items in their orderings, with n chosen by the policy maker based on what their resources can bear. The use of powerful computers, parallel processing, heuristics etc may help increase n.

We use the aforementioned course preferences data from Budish and Cantillon (2012) to simulate the outcome of the revision draft. We find that these outcomes compare favorably with random serial dictatorships and simple drafts with strategic agents - the average rank of the classes in assigned bundles is lower, the distribution of such ranks is more compressed, and the worst off students do better in our model.

Over the past decade several papers have attempted to lay out similar algorithms for solving this problem, but typically rely on the assumption of a large market to secure strategy-proofness as a property. We provide a counter example to show how these algorithms fail to prevent manipulations in a small market, and evidence that they fail to prevent manipulations in the kind of real world markets in which they might be deployed.

While this essay will use the language of students and classes, it should be noted that such mechanisms can be applied to any setting where multiple objects

[3]Columbia's 14 page guide for MBA students can be found here.
[4]No agent can do better than by telling the truth.
[5]In the sense of being Pareto optimal - there are no allocations that make one or more agents better off without making other agents worse off.

must be assigned to agents with heterogenous preferences such as assigning shifts to workers, space in events to conference attendees, players to sports teams, and slots in airports to airlines, among others.

2 Related Literature

The multi-unit assignment problem is well-known and much-studied in mechanism design. We examine this problem from a pragmatic approach---in addition to considering traditionally desirable properties (such as fairness, ex post efficiency, and strategy-proofness), we seek to define an algorithm that incorporates the underlying types of strategic manipulations students engage in. The motivation is to reduce the relative advantage "strategic" players have over "honest" reporters. Our approach levels the playing field by making all players strategic.

Many of the most prominent results in the multi-unit assignment literature are impossibility results. Classic works, like Sönmez (1999), Konishi, Quint, and Wako (2001), among others, highlight the specific difficulties of exporting positive results found in the single-object model to the multi-unit assignment problem. Pápai (2001) demonstrates the only Pareto efficient, non-bossy, and strategy-proof mechanism is a sequential dictatorship. Ehlers and Klaus (2003) extends this by proving dictatorships are the only class of mechanism to satisfy efficiency and coalition strategy-proofness. Despite these findings, schools allotting courses tend not to utilize pure sequential dictatorships, instead often opting for a simple draft or auction mechanism, which are regarded as being fairer.[6]

In this work, we take the underlying insights from Budish and Cantillon (2012)'s proxy draft and apply them to a discrete problem.[7] We find that many of the desirable qualities of their definition of the proxy draft do not uniquely identify a corresponding discrete-world mechanism or algorithm.[8] Although individually distinct from each other, several specific algorithms proposed in the computer science and economics literature satisfy the necessary properties of a proxy draft, despite operating quite differently (see Kominers, Ruberry, and Ullman (2010) and Hoshino and Raible-Clark (2014) for two such examples).

In certain respects, our proposal is a "middle road" between the mechanisms proposed by Budish and Cantillon (2012) and Kominers, Ruberry, and Ullman (2010). Like both, the revision draft relies on a proxy to strategically manipulate individual preferences, thus avoiding the incentive for individuals to attempt these manipula-

[6]Even in cases where schools use first come first served mechanism, it can deviate from pure dictatorship mechanisms in important ways, since eagerness to enroll in a course can be argued to be a stand-in for utility.

[7]Due to the difficulties of working with small markets, many multi-unit assignment problems focus on results derived in a continuum setting (in addition to previously cited works, see Azevedo and Budish (2018).

[8]Specifically, Budish and Cantillon propose a mechanism based on a draft where the central authority: 1) is allowed to "strategize" on behalf of agents by misreporting preferences, and 2) does so with full knowledge of the draft order lottery's realization.

tions themselves. However, our mechanism does not assume that the point of time at which a student is assigned a class is exogenous. We find that the relaxation of this assumption improves the ability of the proxy to optimally manipulate reported preferences. While that means our algorithm occasionally does not terminate, we show the importance of endogenizing run-out times when strategizing for agents.

We also add to the large and growing literature that seeks to compare the welfare implications of different mechanisms. Diebold et al. (2014) use data from two field experiments to compare the the effect of assigning courses via student optimal stable matching, an efficiency adjusted deferred acceptance procedure, and a first come first serve mechanism, finding the two deferred acceptance algorithms tended to outperform the first come first serve alternative. Krishna and Ünver (2008) similarly show the empirical advantages of a Gale-Shapley Pareto-dominant mechanism over auctions. This also builds on the robust body of work experimentally examining various allocation mechanisms.[9] In a slightly different line, Benson, Johnson, and Lybecker (2013) use an instrumental variable regression to estimate undergraduate student bid functions. Related are analyses that examine efficiency considerations in other environments, such as Pathak and Sönmez (2008) and Wu and Zhong (2014), and works that empirically test the observed manipulability of different mechanisms as they exist in practice, such as Pathak and Sönmez (2013) and Agarwal and Somaini (2018). Others - such as Carvalho, Magnac, and Xiong (2019) - highlight the importance of both strategic and distributional impacts of changing allocation mechanisms.

Finally, we argue that our mechanism poses one additional advantage over some of the alternatives: it is more easily understood by agents, which can itself bolster legitimacy. It requires that students submit only an ordering of classes, rather than engaging in a multi-day auction where they must parse the current and historical prices of classes.

The importance of a mechanism or administrative procedure being transparent and easily understood is observed in many administrative settings. Transparency and understandable information transmission has been cited as a necessary ingredient for effective governance e.g. Douglas and A. Meijer (2016). Moreover, some experimental evidence suggests transparency is most useful when subjects have little personal knowledge about the area of administration[10] - such as a course allocation mechanism that students interact with once or twice before graduating. In an experimental setting comparing the Competitive Equilibrium from Equal Incomes (CEEI) and auction mechanisms, Budish and Kessler (2015) report the largest downside of CEEI was a perceived lack of mechanism transparency as reported by participants.

[9] See Featherstone and Niederle (2008), Chen and Sönmez (2006), Kagel and Levin (2001), Engelmann and Grimm (2006), Préget and Thoyer (2014), and Andreoni, Che, and Kim (2007) for various experimental approaches

[10] See Grimmelikhuijsen and A. J. Meijer (2014)

3 Environment

In this section we describe the environment of this problem. For this setup, we are heavily indebted to prior work by Budish (2011) and Hatfield (2009).

Students There are a set of N students $S = \{s_1, ..., s_i, ...s_N\}$ at least two of whom wish to consume more than 2 courses, and all of whom can consume a maximum of Q courses. Note that we are assuming that the number of students is finite.

Courses There are M courses $C = \{c_1, ..., c_j, ..., c_M\}$. Each course has some capacity described by the vector $q = (q_1, ..., q_k, ..., q_M)$ which describes the maximum number of students that can take the course.

Let $\mathbb{C}$ be the powerset of C. We can construct some $\mathbb{S} \subset \mathbb{C}$ that includes all sets of classes that are universally acknowledged as substitutes for each e.g. they occur at the same time, or are sections of the same course. Note that we are assuming the number of courses is finite.

Schedule A schedule σ is a $N \times M$ binary matrix where $x_{ij} = 1$ in row i and column j indicates that student that student i has been allocated course j and $x_{ij} = 0$ indicates otherwise.

A *feasible schedule* has two properties. First, no class is at excess capacity i.e. $\Sigma_i x_{ij} \leq q_j \forall j$. Second, no student is taking more than Q courses i.e. $\Sigma_j x_{ij} \leq Q$. The set of all feasible schedules is given by Ω. j

Preferences Each student has a a von-Neumann Morgenstern utility function u_i : $\Omega \to \mathbb{R}_+$. There are no peer effects, students care only about their own allocation, and perfectly know their own preferences. There preferences are private information.

Monotonicity In this context, a set of preferences is said to have monotonicity if, given an agent i with complete, reflexive and transitive preferences $x_1 \succeq_i \ldots \succeq_i x_j \succeq_i x_M$, and for an arbitrary allocation $C_i \subset C$, $\{C_i, x\} \succeq_i \{C_i, y\} \iff x \succeq_i y$. Consistent with the the literature on multiunit allocation, we will treat preferences as though they were lexicographic, though this is not strictly necessary for our conclusions to hold.

Reports Student i reports their preferences via an ordering $o_i : c_{i1} \succ \ldots \succ c_{il} \succ \ldots \succ c_{iM}$ and where $c_{il} \in \{C, \phi\}$. $c_{il} = \phi$ indicates the cutoff below which the student does not wish to be allocated any courses. A set of orderings is given by $O = \{o_1, \ldots, o_i, \ldots, o_N\}$. The set of all possible orderings is $\mathbf{O}$.

Mechanism The mechanism is a function $f : \mathbf{O} \to \Omega$ that creates a feasible schedule based on student's reports. Thanks to the Revelation Principle, we know we need only consider mechanisms where agents submit their preferences directly.

4 Data

Later in this paper we will be comparing the outcomes of various mechanisms to our preferred mechanism. To do so, we simulate outcomes using the dataset published by Budish and Cantillon (2012). It covers the allocation of second-year elective courses to second-year MBA students at Harvard Business School in 2005–2006. Students choose up to ten elective courses, five per semester, and the full allocation is done at the same time.

The data include a survey of students five most preferred courses, a trial run, and the actual rankings submitted for the allocation process. Budish and Cantillon construct a dataset of true preferences which they use to compare with students' stated preferences, and we are indebted to them for making these data publicly available.

The data cover 456 students choosing from a menu of 67 subjects, with capacities ranging from 12 to 401. Students ranked up to 30 courses. Since our focus is on comparing multiple possible systems to each other, we made a few changes to the process, such as restricting the maximum number of classes to 8, and treating all courses as if they were full courses rather than half-courses, which some were. Simulations were run using Python and analysis was done using STATA; the code is available on request.

5 Revision Draft Mechanism

Students won't try to game the system - by misrepresenting their preferences - if the system strategizes for them. In this section, we lay out a step-by-step mechanism in which the system does so. Note that $f^S(O)$ is the outcome of a simple draft given ordering O. Let L be a set of loop generating orderings.

Step 1: Each student i submits their ordering o_i. An initial set of orderings $O^1 = \{\bigcup_i o_i\}$ is constructed.

Step 2: Each student i randomly draws, without replacement, a number $r_i \in \{1, ..., N\}$.

Step 3: Round $\rho = 1$ starts.

Step 4: At the beginning of Round 1 and all odd numbered rounds, a students' rank is determined by $R_i = r_i$. At the beginning of Round 2 and all even numbered rounds, a students' rank is determined by $R_i = N - r_i + 1$.

Step 5 (simple draft): The highest ranking student is allocated their top available choice, provided the resulting schedule is feasible. The second highest ranking student is allocated their top choice from the remaining options, provided the resulting schedule is feasible, and so on, until the round ends or the cycle terminates.

A round ends when the student with rank N is assigned a course. When a round ends, round $\rho + 1$ starts, and the algorithm returns to Step 4.

A cycle is a set of rounds, and is said to terminate if all courses are at capacity, or if all students are taking Q courses, or if all students do not wish to be assigned any courses among those available. If the cycle terminates, the algorithm goes to Step 6.

Step 6: We start with the highest ranked student, $r_i = 1$

Step 7: For student i, for each possible ordering o_i', the algorithm checks whether they would receives a strictly preferred allocation $f^S(o_i', O_{-i}^y) \succ f^S(o_i, O_{-i}^y)$. If no such ordering is found, we repeat this process for the next highest ranked student.

If one or more such orderings are found for a student, a new set of orderings is constructed $O^{y+1} = \{O_{-i}^y \bigcup o_i^*\}$ where o_i^* is the highest such payoff such that $O^{y+1} \notin L$.

A simple draft is conducted and students are assigned allocation $f^S(O^{y+1})$. If $O^{y+1} = O^i$ for any $i = \{1, ..., y-1\}$ then a loop has been discovered and the algorithm goes to step 8. Otherwise, it returns to the beginning of step 7, analyzing the next highest ranking student, using O^{y+1} instead of O^y.

After the lowest ranked student has been analyzed, if no such ordering is found for a student of *any* rank the algorithm terminates. It is said to *terminate.*

If any such ordering has been found, the algorithm returns to Step 3. *(While retaining the most recent O^y.)*

Step 8 (loops): Consider the ordered set of orderings $\Omega = [O^1, O^2, ..., O^p]$

where $O^p = O^k$ for some $k \in \{1, 2, ..., p-1\}$. O^{k+1} is added to the set of loop generating orderings L and we return to the beginning of step 7, using O^k.

6 Properties

6.1 Convergence

Observation: The revision draft might trigger a loop.

An example is provided in the appendix. It is for this reason that the algorithm checks for loops at each stage of the process.

Claim: The revision draft terminates.

Proof: If no loop is encountered, the algorithm terminates.

If a loop is encountered, an ordering that triggers the loop is added to L and is never considered by the algorithm. Thus the algorithm will only consider sequences of orderings that do not trigger a loop, and will terminate.

6.2 Feasibility

Claim: The revision draft generates a feasible schedule.

The proof for this is trivial: a student is only assigned a course if the resulting schedule is feasible. Therefore, the final schedule will be feasible.

6.3 Dictatorship

A mechanism is a sequential dictatorship if there exists at an ordering of agents such that the first receives their Q favorite items, the second receives their Q favorite items from those left, and so on.

Formally, a mechanism is a sequential dictatorship if there exists an agent n_1, and, for each preference profile $o \in \mathbf{O}$, an ordering of the remaining agents

$$n_2(f_{n_1}(o)), ..., n_N(f_{n_1}(o), ..., f_{n_{N-1}(f_{n_1}(o), ..., f_{n_{N-2(.)}}(o)}(o))$$

such that

$$f_{n_1}(o) = \arg \max_{Z \subseteq \mathbb{M}, |Z| = Q} \left\{ \sum_{m \in Z} a^m_{n_1} \right\}$$

$$f_{n_2(f_{n_1}(o))}(o) = \arg \max_{Z \subseteq \mathbb{M} \setminus f_{n_1}(o), |Z| = Q} \left\{ \sum_{m \in Z} a^m_{n_2(f(n_1(o)))} \right\}$$

and recursively for $i = 3, 4, ..., N$

$$f_{n_i(f_{n_1}(o),\dots,f_{n_{i-1})}(o)}(o) = \arg \max_{Z \subseteq \mathbb{M} \setminus \left[f_{n_1}(o) \cup \bigcup_{j=2}^{i-1} f_{n_j(f_{n_1}(o),\dots,f_{n_{j-1}(.)}(o))} \right], |Z|=Q} \left\{ \sum_{m \in Z} a^m_{n_i(f_{n_1}(o),\dots,f_{n_{i-1}(.)}(o))} \right\}$$

Claim: The revision draft is not a sequential dictatorship

Proof: We provide a counter-example to demonstrate that our mechanism is not a sequential dictatorship. Consider a school with four classes $\alpha, \beta, \gamma, \delta$ with two slots each. Students take two classes each. There are four students who have the following preferences:

$$A : \alpha \succ \beta \succ \gamma \succ \delta$$
$$B : \alpha \succ \beta \succ \gamma \succ \delta$$
$$C : \alpha \succ \beta \succ \gamma \succ \delta$$
$$D : \alpha \succ \beta \succ \gamma \succ \delta$$

Regardless of their initial ranks, no student will be assigned both α and β. *Proof ends.*

6.4 Manipulability

A mechanism is *strategy proof* if agents can do no better by misrepresenting their preferences. Formally, a mechanism F is weakly strategy-proof if $f^{RD}(o_i, O_{-i}) \succeq_i f^{RD}(o'_i, O_{-i})$ for all agents $i = \{1, \dots, N\}$, true orderings $o_i \in \mathbf{O}$, and all possible orderings $o'_i \in \mathbf{o}$.

Proposition: The revision draft is strategy proof for profiles that do not ever encounter a loop (i.e. if L is a null set when the algorithm terminates).

Proof: Suppose not. Then there exists an agent i and an ordering $o'_i \in \mathbf{o}$ such that $f^{RD}(o'_i, O_{-i}) \succ f^{RD}(o_i, O_{-i})$ which implies that for some y, $O^y = (o'_i, O_{-i})$ which implies that $f^{RD}(o_i, o_{-i}) \sim f^{RD}(o'_i, O_{-i})$, which violates $f^{RD}(o'_i, O_{-i}) \succ f^{RD}(o_i, O_{-i})$. *Proof ends.*

Conjecture: *If preferences are lexicographic, then revision draft is strategy proof*

Intuition: Let k be the agent whose manipulation starts the loop. Within each loop, there is a set of orderings that gives agent k a less preferred payoff and a set of orderings that would give them a more preferred payoff. If k had lexicographic preferences, they would choose not to enter the loop, which is the strategy that we implement.

6.5 Efficiency

A mechanism f is *Pareto optimal* if there exists no way to make an agent better off without making another agent worse off. Formally, a mechanism f is Pareto optimal if there does not exist an allocation X such that $X \succeq_i f_i(O)$ for all $i = \{1, 2, ..., N\}$ and $X \succ_i f_i(O)$ for some $i = 1, ..., N$.

Proposition: The revision draft is Pareto optimal

Proof: Suppose not. Let R be the allocation generated by the revision draft and S be a superior allocation such that $S \succeq_i R$ for all $i = 1, ..., N$ and $S \succ_i R$ for some $i = 1, ..., N$. Let A be a student for whom $S \succ_A R$. Lexicographic preferences imply that A was allocated at least one course s in S that they were denied in R and allocated at least one course r in R where $s \succ_A r$. Let ω be the set of courses that A was allocated before α. Let $o'_A = \{\omega, s, r, ...\}$. We observe that $f^S(o'_A, O^y_{-A}) \succ f(O^y)$. This violates termination. *Proof ends.*

7 Shortcomings of the Mechanism

This algorithm, implemented as it currently is, searches for a Nash Equilibrium by what computer scientists call brute force. For each student it invokes each possible alternative ordering, and checks whether that ordering improves the students' outcomes. In the HBS dataset, students submitted their rankings of up to 30 classes, which can be ordered in 30! ways i.e. in 2.65×10^{32} ways. Simply computing the permutations would take longer than the expected lifetime of the universe, and this process would have to be repeated for each student, and this entire process must be repeatedly iterated until a Nash Equilibrium is found.

To tackle this problem, we propose the following restriction: allow the algorithm to manipulate only the top n items in a students ordering. Now, the algorithm need only check $n!$ permutations of orderings for each student. This solution has the advantage of being scalable: policy makers can decide what value of n they wish to use based on the nature of the economy they are managing, the computing power they have access to, the programming talent they have available, and so on.

Given the resources available to us, we were able set n as high as 5. We utilize this restriction for the remainder of the paper.

8 Comparisons to Alternatives

8.1 Random Serial Dictatorship

A Random Serial Dictatorship (RSD) is conducted as follows: a lottery for ranks is conducted. The highest ranking student is assigned all of their top choices. The second highest ranking student is assigned all of their choices from what items remain, and so on.

The RSD has many desirable properties: it is Pareto-efficient, strategy-proof, non-dictatorial, and non-bossy. Indeed, Hatfield (2009) shows that it is the *only* mechanism that has these characteristics. The downside is that it is ex-post "unfair" in the sense that it is *callous.* From Budish and Cantillon (2012):

> *Callousness: An anonymous multi-unit random priority mechanism is callous if there exists $n \in \{2, ..., m\}$ and a random priority draw, such that a strictly positive measure of students get their nth choosing time before another set of students get their $(n-1)$th choosing time. Otherwise it is non-callous.*

Callousness is an undesirable property, and the RSD mechanism is extreme example of it - in an economy of two students, one will take all their turns before the other.

To compare the welfare effects of RSD to the revision draft, we simulate them in the HBS economy. The distributional effects of these allocations can be seen in Figures 1 and 2. In Table 1 we see that the average rank in the assigned bundle of a student in the revision draft is 5.8 with a standard deviation of 0.92. The average rank in the assigned bundle of a student in the RSD is 6.4 with a standard deviation of 2.3. The RSD is both worse on average and has considerable more variance in outcomes. Most strikingly, the worst off student in the revision draft acquired a bundle with an average rank of 9.9, while the worst off student in the RSD acquired a bundle with an average rank of 14.6.

8.2 Simple Draft

A simple draft is conducted as follows: a lottery for ranks is conducted, the highest ranking student is allocated their top choice. The second highest ranking student is allocated their top choice from those choices left and so on. When the lowest ranking student is allocated their top choice from remaining options, the ranks are reversed and the process continues until all classes are at capacity or all students are allocated their quota from acceptable choices.

While the resulting allocations are more equal than randomized serial dictatorships, the problem with this method is that it is manipulable. Consider the following economy with three students who must take 2 classes each, and three classes that can accommodate 2 students each:

$$A : \gamma \succ \alpha \succ \beta$$
$$B : \alpha \succ \beta \succ \gamma$$
$$C : \alpha \succ \beta \succ \gamma$$

Regardless of their realized rank, A's best strategy is to report $\alpha \succ \gamma \succ \beta$ as their preferences. There is no chance that γ will full by the time the mechanism takes their second option into account, and these false preferences give A a $2/3$ chance to get αinstead of β.

This is a simple illustration of a general strategy that students can follow - rank popular classes high and unpopular classes low. The simple draft was the method of course allocation at Harvard Business School that was analyzed by Budish and Cantillon (2012) who showed that the mechanism is "manipulable in theory, manipulated in practice, and that these manipulations cause significant welfare loss". Because students are trying to guess the relative popularity of courses while uncertain of their own realized ranks, they sometimes end up losing out on classes they would have liked to take, or are denied popular courses they sincerely ranked highly.

How does our mechanism stack up against the simple draft, after incorporating the manipulations of students? Figures 3 and 4 show the distribution of the average and median ranks respectively. Our draft is has lower ranks on average and lower variance, a conclusion that is echoed in Table 1.

8.3 Proxy Drafts

Proposed solutions to the multiunit allocation problem typically assume a continuum economy, equivalent to treating consumers as price takers. The properties that they have - such as strategy-proofness - do not hold in a small economy. In this section, we describe one such proposal, and show that the mechanism fails to generate a strategy-proof allocation.

Budish and Cantillon (2012) propose a *proxy draft*. Each student s submits their ordering of classes $P_s : c_1, c_2, ...$ After they do, a lottery is held in which

> ... each student draws a random priority number independently from the uniform distribution on $[0, 1]$, and a student who draws priority x gets choosing time x in the first round, $2 - x$ in the second round, $2 + x$ in the third round, etc.

A draft is then conducted. At each time period, the student who has the corresponding choosing time is assigned their top choice of classes among those that are left. At the end of the draft, the computer notes the time at which the last unit of a course is allocated to a student, called the course's runout time. The runout time of course i is t_i^* and is assumed to be exogenous.

Next, the algorithm attempts to find a Nash Equilibrium in which each student s is submitting an ordering P_s that is a best response to the orderings of other students. To do so, the proxy agents order their preferences such that they request a class as late as possible. For each student s, the algorithm makes a list of choosing times $T = \{\tau_1, \tau_2, ..., \tau_m\}$ ordered such that $\tau_1 < \tau_2 < ... < \tau_m$. Starting with the students most preferred option i, the computer checks if there exists *any* choosing time $\tau \in T$ such that $\tau \leq t_i^*$ i.e. whether it is possible for the student to acquire the item at all. If not, it looks at the next option. If so, it finds the largest value of $\tau \in T$ such that $\tau \leq t_i^*$ i.e. the latest time that the student can request course i and still get it. This time is removed from the list of available choosing times, and the corresponding position on the ordering is then set to i

Finally, the draft is then repeated using the new orderings.

In large markets, runout times are indeed exogenous and the new orderings are indeed best responses to the runout times. However, in small markets, by changing their orderings, an agent may change the runout time. The algorithm fails to incorporate this effect. Consider an economy with 4 students, each of whom wants 2 classes, and 4 classes, each of which can accommodate 2 students. Suppose that the students are ranked in the following order:

$$A : \alpha \succ \beta \succ \gamma \succ \delta$$
$$B : \alpha \succ \beta \succ \gamma \succ \delta$$
$$C : \beta \succ \gamma \succ \alpha \succ \delta$$
$$D : \beta \succ \delta \succ \gamma \succ \alpha$$

For clarity, we will use the choosing times $\{1, 2, 3...\}$ to denote when the agents make their choices. We will apply a label to show when each option was allocated, so κ^t means that a unit of κ as allocated at time t to the corresponding agent. Using this notation, we can see the initial draft play out:

$$A : \alpha^1 \succ \beta \succ \gamma \succ \delta^8$$
$$B : \alpha^2 \succ \beta \succ \gamma^7 \succ \delta$$
$$C : \beta^3 \succ \gamma^6 \succ \alpha \succ \delta$$
$$D : \beta^4 \succ \delta^5 \succ \gamma \succ \alpha$$

The runout times are $t_\alpha^* = 2$, $t_\beta^* = 4$, $t_\gamma^* = 7$ and $t_\delta^* = 8$. Note that for agent A, the choosing times are $\{1, 8\}$. The proxy draft will check if A's most preferred option

α is attainable. Since $t^*_\alpha = 2$ which is less than choosing time 1, it is attainable. In the eyes of the algorithm, choosing time 1 is the latest at which A can request $\alpha$and be allocated it, so the top ranked option is set at α and $\tau_A = 1$ is removed from the set of available choosing times. The next two options - β and γ - are not attainable, but the last option δ is attainable at time 8. Thus the proxy draft will rearrange A's orderings to $\alpha \succ \delta \succ \beta \succ \gamma$.

However, A's actual best response is $\beta \succ \alpha \succ \gamma \succ \delta$, which would result in the agent being assigned both their top options. By lowering their ranking of α, A increases it's runout time. By raising their ranking of β, A decreases it's runout time. These changes are not considered possible by the proxy draft algorithm, and so it fails to discover this response. As a result of this, the proxy draft does not generate a Nash Equilibrium, and is not strategy-proof.

It is possible, however, that real-world economies are sufficiently large for this to be an unimportant concern. The course allocation at Harvard Business school is an economy with 456 students and 67 classes. Holding ranks fixed, we ran a proxy draft, and then allowed students to manipulate the ranks of their top $n \in [2, 5]$ choices.[11] The results of this exercise can be found in Table 2. Over 16% of students can improve their allocation by simply reversing the top 2 courses in their allocation. If we allow them to manipulate the top 4 courses, over half of them benefit from deviating from submitting their true preferences. Thus, it is untrue that real-world economies are sufficiently large to make the proxy draft strategy-proof.

8.4 Auctions

Auction mechanisms are one of the most popular ways for business schools to assign courses to students, though little agreement exists between schools about the specifics of auction mechanisms used. For example, the Ross School of Business at the University of Michigan allocates a set number of "points" every term to all students without allowing unused points to be carried over. Kellogg endows first and second year MBA students with different endowments - 2000 "points" for first years, 3000 for second years - and allows carryover *across* quarters, but not across years. Columbia Business School endows all students their first semester with "lifetime bid points" which are carried throughout a student's time in the program. In the event of a tie, most business schools use a randomization device to pick the "winning bidders," while others, like the Foster School of Business, give priority to the chronologically earliest students who input a jointly-winning bid.[12] MIT Sloan utilizes a hybrid priority-bidding mechanism, where seats are allocated first by a student's priority ranking (with higher priority given to Sloan students, and those

[11] $n \geq 6$ were not considered due to limitations in the computing power available to us.

[12] And hence incorporating an element of a first come first served within the auction mechanism framework.

in later years), with ties being broken by student bids.

The wide variety of auction mechanisms makes broad, generalizable claims difficult to conclusively demonstrate. Mechanisms that incorporate stricter priority tiers reduce manipulability at the cost of reduced fairness. Schools that allow saving "points" across instructional terms change bidding strategies in ways with unclear impacts on student welfare.[13] Many course auctions only require payment for classes that are over-enrolled, using the auction as a tiebreaker. However this can lead to significant discontinuities in course price based on relatively arbitrary characteristics.[14]

Although relatively straightforward to implement, auctions have been shown to impose potentially significant losses to student welfare. Course prices can change significantly between semesters,[15] leading to sub-optimal course scheduling by students. Many of these issues are described in great detail by Sönmez and Ünver (2010). By assuming students behave as price takers, the literature discounts one of the main ways "unintentional" strategic behavior is likely to impact course prices (and thus student welfare). Those who submit bids to get a class "if the price is right" can push up the market clearing price, decreasing the auction budget for students who ultimately win a seat in the course. This has potentially large effects in circumstances where individuals know each other and have repeated interactions.

Nevertheless, there are some dimensions against which we can compare our proposed mechanism. First, some experimental results have suggested sizable efficiency costs for auctions in a course-allocation setting e.g. Krishna and Ünver (2008) and Budish and Kessler (2015). Beyond these direct efficiency costs, there is implicit evidence of attention and information costs associated with course auctions - the aforementioned business schools' auction procedures all span more than a week; the reason given by many is the need for students to familiarize themselves with historical course prices, determine their optimal strategies, and update bids after early rounds of bidding have concluded. Second, we return to the initial concern of the fairness implications of manipulable mechanisms. Others have noted the asymmetry inherent in manipulation of an auction mechanism: some students are likely to be more able to effectively manipulate bids in order to enhance their own

[13] Limiting budgets to a single semester or quarter incentivize students to mix high and low demand courses in order to exhaust their allotments. This could be efficiency improving if it functioned as a coordination mechanism. On the other hand, this could easily lead to welfare losses if student interests change during their time at the university, if many highly demanded courses have prerequisites, or if courses are offered somewhat unpredictably.

[14] Consider a class with q seats that is highly sought after by a subgroup of students of size $q + 1$. The expected outcome of the auction procedure would be a high price for the class, which substantially impacts the other classes the auction winners can afford. However, if one of the $q + 1$ students has a scheduling conflict which precluded them from being able to take the class, or if the school had assigned the course to a larger classroom, the price of the course would be 0. These features are not necessarily observable by students, which can lead to inefficiencies, even in expectation.

[15] This has been documented as far back as Graves, Schrage, and Sankaran (1993)

payoff at the expense of others.[16] Our proposed mechanism reduces this unfairness by putting all manipulation power in the hands of the centralized proxy.

8.5 Approximate Competitive Equilibrium for Equal Outcomes

One alternative approach to the multi-unit allocation problem is to utilize the *Competitive Equilibrium for Equal Incomes* (CEEI) solution, based off Varian (1974). As Budish (2011) shows, although the CEEI solution to a course assignment problem may not exist. perturbing incomes slightly around equality is sufficient to guarantee the existence of an equilibrium allocation. This refinement, dubbed the *Approximate Competitive Equilibrium for Equal Outcomes* (A-CEEI), is then applied to the course registration system at the Wharton School of Business, replacing an auction procedure. Despite some problems surrounding agent reporting and difficulty understanding the mechanism described in Budish and Kessler (2015), Budish, Cachon, et al. (2017) demonstrate efficiency gains from the A-CEEI solution (through an application called "Course Match") over an auction mechanism at Wharton. The development of A-CEEI has led to other modifications of CEEI-type solutions to allocation problems, such as Aziz (2015).

The A-CEEI mechanism, as Othman, Sandholm, and Budish (2010) succinctly describe, utilizes a 'behind the scenes' auction mechanism that is entirely conducted by the central authority. First, students report preferences over classes.[17] Second, the mechanism randomly determines budgets for students, limited to a particular (narrow) range. Third, given student preferences and budgets, the mechanism simulates an auction procedure, finds the market clearing prices, and assigns schedules.

Based on the results in Budish, Cachon, et al. (2017), there is little question the A-CEEI mechanism, at least as applied to course allocation at Wharton, improved student outcomes on average compared to an auction across a number of welfare metrics, and yielded higher student satisfaction based on survey results. At the same time, Budish and Kessler (2015) note a small but significant subset of an experimental population did not understand the mechanism and, as a result, reported preferences sub-optimally. This confusion should not be ignored - although to a lesser degree, this reintroduces a similar asymmetry as was present in auction mechanisms.[18] We argue that a revision draft is better able to alleviate deadweight losses associated with confusion: not only is the preference reporting mechanism

[16]See this newspaper article, for example, which suggests certain students - they highlight former financial traders - as being more effective with bid manipulation.

[17]Theoretically, students would report preferences over every possible class schedule. Due to the logistical impossibility of this task, Budish and Kessler (2015) demonstrate an alternative method for gathering sufficient preference information. Students are ask for course rankings, relative utility gained from courses (in addition to their ranking) and about a limited number (5 or 10) of specific schedules, presented as a series of pairwise comparisons.

[18]In an auction, a student "competing against" a former securities trader was at a strategic disadvantage. Here, a student who cannot understand the reporting system (with three related, yet distinct elements) is harmed.

simpler - as students only report course orderings - the functional operation of the mechanism can be more intuitively explained and understood.

9 Discussion & Conclusions

The multiunit allocation problem is a deep and difficult one. Existing methods of solving it typically suffer from at least one of three problems: they are unfair, or manipulable, or confusing. The revision draft overcomes these issues.

First-come-first-served and random serial dictatorships are simple and not manipulable, but suffer from generating allocations that are highly unequal - some students get all their preferred choices while others get none. They are especially awful for students who end up at the bottom of their ranking, or who are last to sign up - in the real-world economy we analyzed, this resulted in them getting, on average, their 15th choice. We show that the revision draft generates allocations that are far more equal, while still being strategy-proof.

Simple drafts are easy to understand and result in more equal allocations, but are easy to manipulate. As a result of such manipulations, the welfare of students is lower (and the inequality higher) than it would be under revision drafts, which is a strategy proof mechanism. A similar argument applies to methods such as proxy drafts that require large markets to prevent manipulation - real world markets are not sufficient for this purpose.

Auctions seem fair, strategy-proof, and simple in theory, but are actually manipulable in small economies, and take a tremendous amount of time to implement. The prices in any given year are often disconnected from prices in previous years. Methods such as A-CEEI attempt to overcome these problems, but the opaque nature of process undermines its legibility. Revision drafts, by way of contrast, are easy to understand and implement, and require very little input from students - either a ranking of classes, or a ranking of classes supplemented by some measure of the utility they derive from them.

Nonetheless, the revision draft is not a perfect solution. It requires a great deal of computational resources, to the point that executing it as much a problem of computer science as it is of economics.

There is a great deal of work yet to be done in this space, such as relaxing the monotonicity assumption to incorporate substitutes and complements. Furthermore, most draft-based mechanisms reverse ranks in subsequent rounds, but there is no reasons to believe that this is the optimal method of balancing ranks.

References

Agarwal, Nikhil and Paulo Somaini (2018). "Demand analysis using strategic reports: An application to a school choice mechanism". In: *Econometrica* 86.2, pp. 391–444.

Andreoni, James, Yeon-Koo Che, and Jinwoo Kim (2007). "Asymmetric information about rivals' types in standard auctions: An experiment". In: *Games and Economic Behavior* 59.2, pp. 240–259. ISSN: 0899-8256. DOI: https://doi.org/10.1016/j.geb.2006.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0899825606001217.

Azevedo, Eduardo M and Eric Budish (2018). "Strategy-proofness in the Large". In: *The Review of Economic Studies*. DOI: 10.1093/restud/rdy042. URL: https://doi.org/10.1093/restud/rdy042.

Aziz, Haris (2015). "Competitive equilibrium with equal incomes for allocation of indivisible objects". In: *Operations Research Letters* 43.6, pp. 622–624. ISSN: 0167-6377. DOI: https://doi.org/10.1016/j.orl.2015.10.001. URL: https://www.sciencedirect.com/science/article/pii/S0167637715001327.

Benson, Cassandra Michele, Daniel K. N. Johnson, and Kristina M. Lybecker (2013). "Bidding for Classes: Course Allocation Under the Colorado College Auction System". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2257921. URL: https://doi.org/10.2139/ssrn.2257921.

Budish, Eric (2011). "The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes". In: *Journal of Political Economy* 119.6, pp. 1061–1103. DOI: 10.1086/664613. URL: https://ideas.repec.org/a/ucp/jpolec/doi10.1086-664613.html.

Budish, Eric, Gérard P. Cachon, et al. (2017). "Course Match: A Large-Scale Implementation of Approximate Competitive Equilibrium from Equal Incomes for Combinatorial Allocation". In: *Oper. Res.* 65.2, 314?336. ISSN: 0030-364X. DOI: 10.1287/opre.2016.1544. URL: https://doi.org/10.1287/opre.2016.1544.

Budish, Eric and Estelle Cantillon (2012). "The Multi-unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard". In: *American Economic Review* 102.5, pp. 2237–71. DOI: 10.1257/aer.102.5.2237. URL: https://www.aeaweb.org/articles?id=10.1257/aer.102.5.2237.

Budish, Eric and J. Kessler (2015). "Experiments as a Bridge from Market Design Theory to Market Design Practice: Changing the Course Allocation Mechanism at Wharton". In:

Carvalho, José Raimundo, Thierry Magnac, and Qizhou Xiong (2019). "College choice, selection, and allocation mechanisms: A structural empirical analysis". In: *Quantitative Economics* 10.3, pp. 1233–1277. DOI: https://doi.org/10.3982/QE951. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE951. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/QE951.

Chen, Yan and Tayfun Sönmez (2006). "School choice: an experimental study". In: *Journal of Economic Theory* 127.1, pp. 202–231. ISSN: 0022-0531. DOI: https://doi.org/10.1016/j.jet.2004.10.006. URL: https://www.sciencedirect.com/science/article/pii/S0022053104002418.

Diebold, Franz et al. (2014). "Course Allocation via Stable Matching". In: *Business & Information Systems Engineering* 6.2, pp. 97–110. DOI: 10.1007/s12599-014-0316-6. URL: https://doi.org/10.1007/s12599-014-0316-6.

Douglas, Scott and Albert Meijer (2016). "Transparency and Public Value—Analyzing the Transparency Practices and Value Creation of Public Utilities". In: *International Journal of Public Administration* 39.12, pp. 940–951. DOI: 10.1080/01900692.2015.1064133. URL: https://doi.org/10.1080/01900692.2015.1064133.

Ehlers, Lars and Bettina Klaus (2003). "Coalitional strategy-proof and resource-monotonic solutions for multiple assignment problems". In: *Social Choice and Welfare* 21.2, pp. 265–280. ISSN: 01761714, 1432217X. URL: http://www.jstor.org/stable/41106560.

Engelmann, Dirk and Veronika Grimm (2006). *Bidding Behavior in Multi-Unit Auctions - An Experimental Investigation*. Working Paper Series in Economics 24. University of Cologne, Department of Economics. URL: https://ideas.repec.org/p/kls/series/0024.html.

Featherstone, Clayton and Muriel Niederle (2008). *Ex Ante Efficiency in School Choice Mechanisms: An Experimental Investigation*. NBER Working Papers 14618. National Bureau of Economic Research, Inc. URL: https://ideas.repec.org/p/nbr/nberwo/14618.html.

Graves, Robert L., Linus Schrage, and Jayaram Sankaran (1993). "An Auction Method for Course Registration". In: *Interfaces* 23.5, pp. 81–92. ISSN: 00922102, 1526551X. URL: http://www.jstor.org/stable/25061804.

Grimmelikhuijsen, Stephan G and Albert J Meijer (2014). "Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment". In: *Journal of Public Administration Research and Theory* 24.1, pp. 137–157.

Hatfield, John William (2009). "Strategy-proof, efficient, and nonbossy quota allocations". In: *Social Choice and Welfare* 33.3, pp. 505–515.

Hoshino, Richard and Caleb Raible-Clark (2014). "The Quest Draft: An Automated Course Allocation Algorithm". In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. AAAI'14. Québec City, Québec, Canada: AAAI Press, 2906?2913.

Kagel, John H. and Dan Levin (2001). "Behavior in Multi-Unit Demand Auctions: Experiments with Uniform Price and Dynamic Vickrey Auctions". In: *Econometrica* 69.2, pp. 413–454. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2692237.

Kominers, Scott Duke, Mike Ruberry, and Jonathan Ullman (2010). "Course Allocation by Proxy Auction". In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 551–558. DOI: 10.1007/978-3-642-17572-5_49. URL: https://doi.org/10.1007/978-3-642-17572-5_49.

Konishi, Hideo, Thomas Quint, and Jun Wako (2001). "On the Shapley?Scarf economy: the case of multiple types of indivisible goods". In: *Journal of Mathematical Economics* 35.1, pp. 1–15. ISSN: 0304-4068. DOI: https://doi.org/10.1016/S0304-4068(00)00061-6. URL: https://www.sciencedirect.com/science/article/pii/S0304406800000616.

Krishna, Aradhna and M. Utku Ünver (2008). "Improving the Efficiency of Course Bidding at Business Schools: Field and Laboratory Studies". In: *Marketing Science* 27.2, pp. 262–282. ISSN: 07322399, 1526548X. URL: http://www.jstor.org/stable/40057101.

McKenzie, Lindsay (2019). "A Class Registration Bot Backfires". In: *Inside Higher Ed*. URL: https://www.insidehighered.com/news/2019/01/15/computer-program-automatically-registers-students-classes-has-unintended.

Othman, Abraham, Tuomas Sandholm, and Eric Budish (2010). "Finding Approximate Competitive Equilibria: Efficient and Fair Course Allocation". In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1*. AAMAS '10. Toronto, Canada: International Foundation for Autonomous Agents and Multiagent Systems, 873?880. ISBN: 9780982657119.

Pápai, Szilvia (2001). "Strategyproof and Nonbossy Multiple Assignments". In: *Journal of Public Economic Theory* 3.3, pp. 257–271. DOI: https://doi.org/10.1111/1097-3923.00066. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1097-3923.00066. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1097-3923.00066.

Pathak, Parag A. and Tayfun Sönmez (2008). "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism". In: *The American Economic Review* 98.4, pp. 1636–1652. ISSN: 00028282. URL: http://www.jstor.org/stable/29730139.

— (2013). "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation". In: *American Economic Review* 103.1, pp. 80–106. DOI: 10.1257/aer.103.1.80. URL: https://www.aeaweb.org/articles?id=10.1257/aer.103.1.80.

Préget, Raphaële and Sophie Thoyer (2014). "Does the competition structure impact the performance of multi-unit auctions ? An experimental investigation". In: *Recherches économiques de Louvain* 80.2, p. 85. DOI: 10.3917/rel.802.0085. URL: https://doi.org/10.3917/rel.802.0085.

Sönmez, Tayfun (1999). "Strategy-proofness and Essentially Single-valued Cores". In: *Econometrica* 67.3, pp. 677–689. DOI: https://doi.org/10.1111/1468-0262.00044. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0262.00044. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00044.

Sönmez, Tayfun and M. Utku Ünver (2010). "Course Bidding at Business Schools". In: *International Economic Review* 51.1, pp. 99–123. ISSN: 00206598, 14682354. URL: http://www.jstor.org/stable/25621516.

Varian, Hal R (1974). "Equity, envy, and efficiency". In: *Journal of Economic Theory* 9.1, pp. 63–91. ISSN: 0022-0531. DOI: https://doi.org/10.1016/0022-0531(74)90075-1. URL: https://www.sciencedirect.com/science/article/pii/0022053174900751.

Wu, Binzhen and Xiaohan Zhong (2014). "Matching mechanisms and matching quality: Evidence from a top university in China". In: *Games and Economic Behavior* 84, pp. 196–215.

A Loops

Consider a problem with 4 students A, B, C, D, and 4 classes $\alpha, \beta, \gamma, \delta$, each student can be enrolled in a maximum of 2 classes, and each class can enroll a maximum of 2 students. They are ranked in the order A, B, C, D. The preferences are as follows, including the allocation after the first simple draft O^1:

$$A : \underline{\alpha} \succ \beta \succ \gamma \succ \underline{\delta}$$
$$B : \underline{\alpha} \succ \beta \succ \underline{\gamma} \succ \delta$$
$$C : \underline{\beta} \succ \underline{\gamma} \succ \alpha \succ \delta$$
$$D : \underline{\beta} \succ \underline{\delta} \succ \alpha \succ \gamma$$

This outcome will prompt A to change their orderings to generate O^2:

$$A' : \underline{\gamma} \succ \underline{\alpha} \succ \beta \succ \delta$$
$$B : \underline{\alpha} \succ \beta \succ \gamma \succ \underline{\delta}$$
$$C : \underline{\beta} \succ \underline{\gamma} \succ \alpha \succ \delta$$
$$D : \underline{\beta} \succ \underline{\delta} \succ \alpha \succ \gamma$$

This outcome will prompt B to change their orderings to generate O^3:

$$A' : \underline{\gamma} \succ \alpha \succ \beta \succ \underline{\delta}$$
$$B' : \underline{\beta} \succ \underline{\alpha} \succ \gamma \succ \delta$$
$$C : \underline{\beta} \succ \underline{\gamma} \succ \alpha \succ \delta$$
$$D : \beta \succ \underline{\delta} \succ \underline{\alpha} \succ \gamma$$

This outcome will prompt A to change their orderings to generate O^4:

$$A'' : \underline{\alpha} \succ \underline{\gamma} \succ \beta \succ \delta$$
$$B' : \underline{\beta} \succ \alpha \succ \underline{\gamma} \succ \delta$$
$$C : \underline{\beta} \succ \underline{\gamma} \succ \alpha \succ \delta$$
$$D : \beta \succ \underline{\delta} \succ \underline{\alpha} \succ \gamma$$

This outcome will prompt B to revert to their original ordering to generate O^5:

$$\begin{aligned} A'' &: \underline{\alpha} \succ \gamma \succ \beta \succ \underline{\delta} \\ B &: \underline{\alpha} \succ \beta \succ \underline{\gamma} \succ \delta \\ C &: \underline{\beta} \succ \underline{\gamma} \succ \alpha \succ \delta \\ D &: \underline{\beta} \succ \underline{\delta} \succ \alpha \succ \gamma \end{aligned}$$

This will prompt A to revert to the ordering that generates O^2, establishing a loop. Note that this process can be thought of as A and B fighting over who is assigned the more desirable class γ.

In Step 7 of our algorithm, we note that the set of people who never changed their orderings in the loop is C and D. They are guaranteed their allocations in the first ordering that is part of the loop O^2. The only orderings that generate this allocation are O^2 and O^5. The agents are randomly allocated either $f^S(O^2)$ or $f^S(O^5)$.

B Tables and Figures

Table 1: Mean Ranks of Bundles Generated by Various Mechanisms

Mechanism	Mean	SD	Maximum
Simple Draft with True Preferences	5.57	0.67	8.00
Simple Draft with Strategic Preferences	6.59	1.21	11.25
Random Serial Dictatorship	6.35	2.30	14.62
Revision Draft	5.76	0.92	9.88
Proxy Draft	5.58	0.75	8.25

Notes: Data about true prefences, stated preferences, and class capacities are from Budish and Cantillon (2012), which analyzes course allocation for a cohort of the Harvard Business school. The data included 456 students and 96 classes. Simulations of these mechanisms were run on Python and analysed in STATA. The revision draft simulation limited students to manipulating their top 5 classes. Code is avaible on request.

Table 2: Improvements After Proxy Draft

n	Improvements
2	76
3	143
4	250
5	291

Notes: Fixing the ranks, students were permitted tomanipulate the top *n* objects in their true orderings. "Improvments" refers to the number of improvements found.

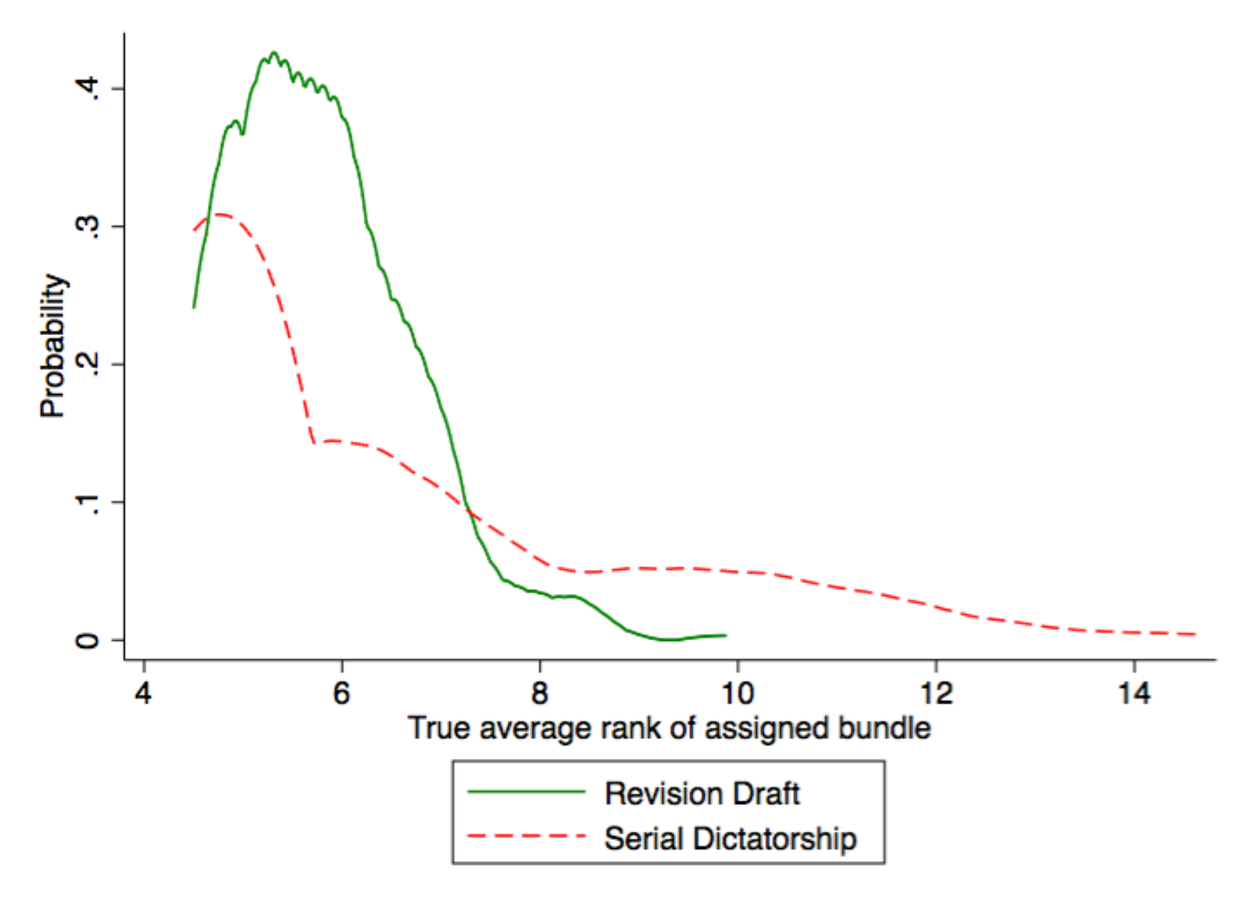

Figure 1: Distribution of the Average Rank Received: Random Serial Dictatorship versus Revision Draft

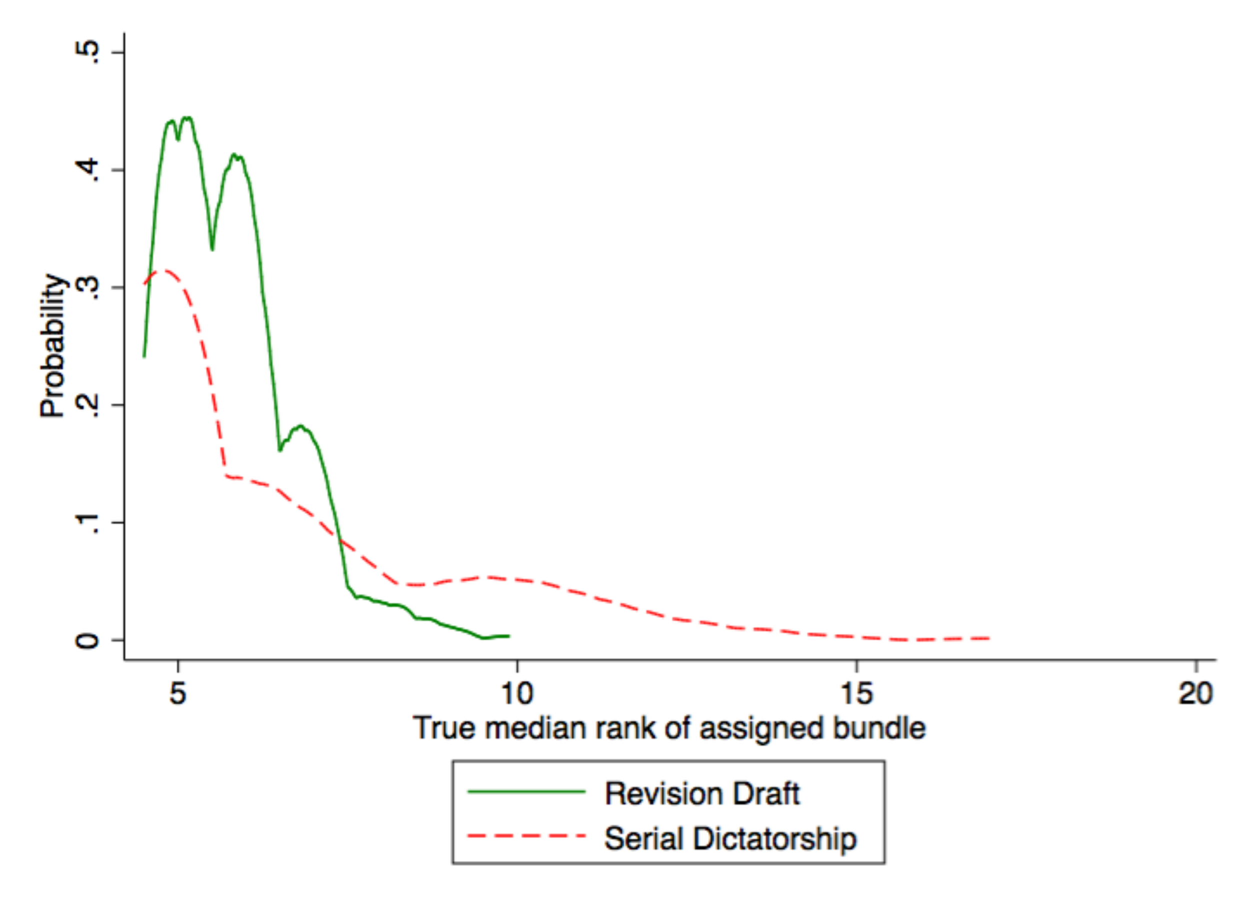

Figure 2: Distribution of the Median Rank Received: Random Serial Dictatorship versus Revision Draft

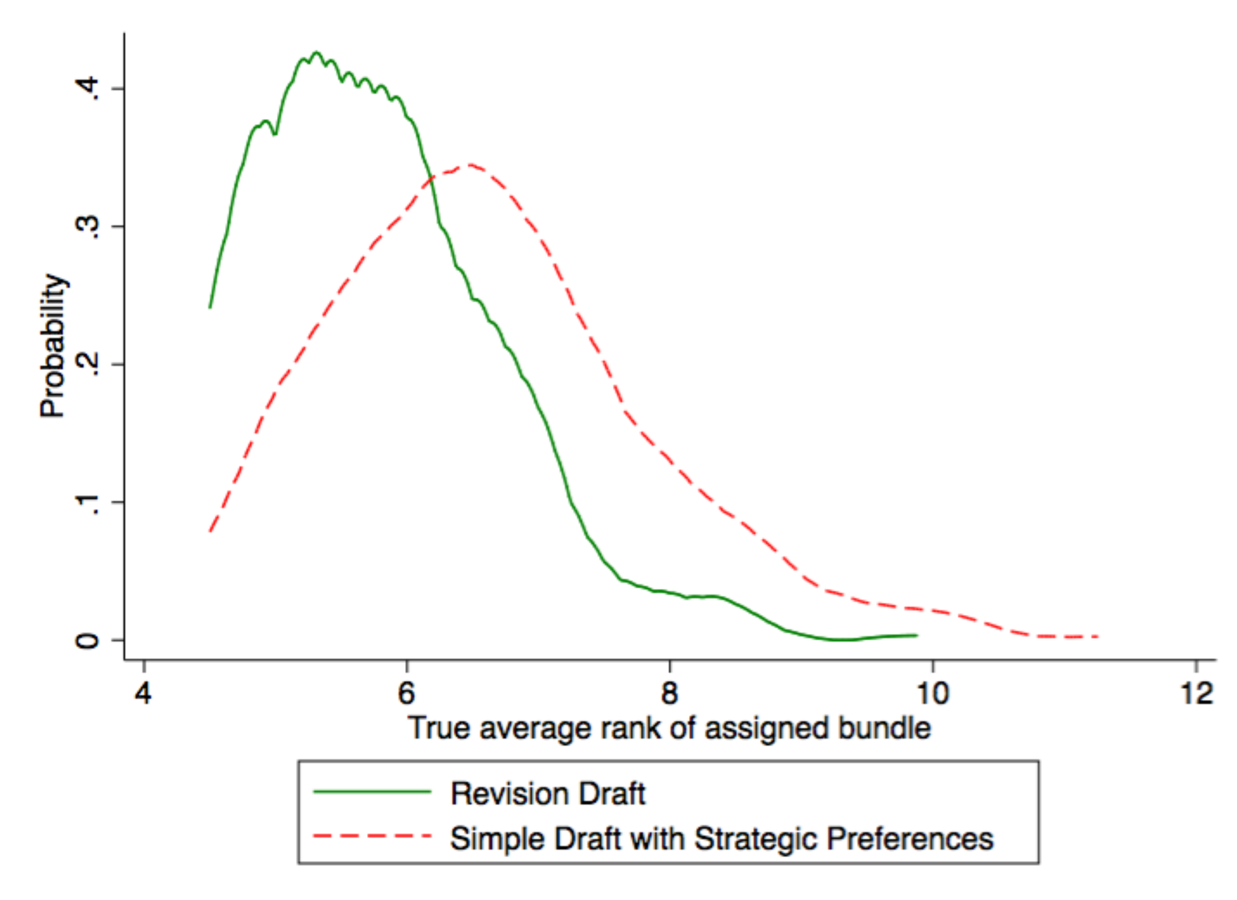

Figure 3: Distribution of the Average Rank Received: Simple Draft versus Revision Draft

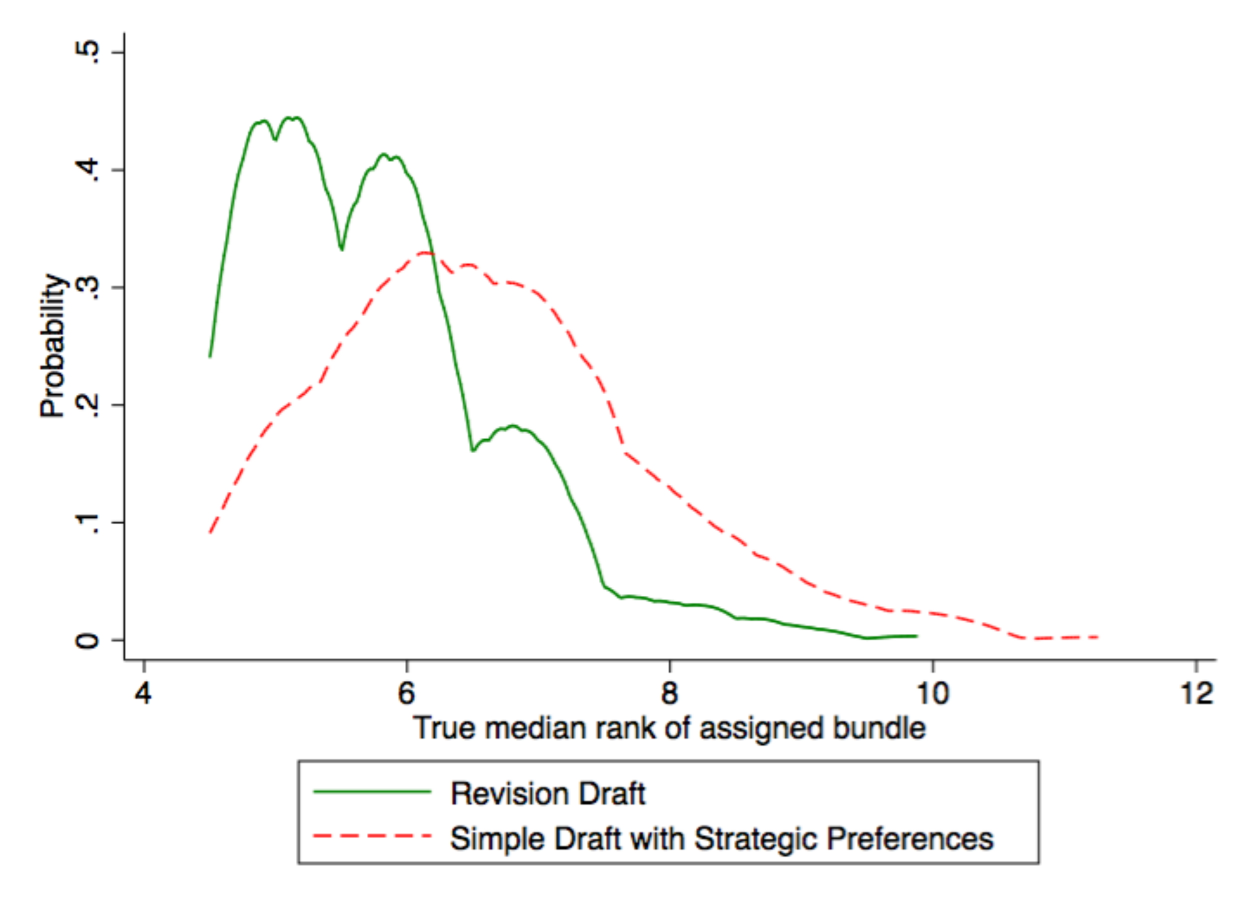

Figure 4: Distribution of the Median Rank Received: Simple Draft versus Revision Draft