Title
Essays in Peer Effects and the Economics of Crime

Author
Stevenson, Megan

Publication Date
2016

Peer reviewed|Thesis/dissertation

Essays in Peer Effects and the Economics of Crime

by

Megan Stevenson

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Agricultural and Resource Economics
in the
Graduate Division
of the
University of California, Berkeley.

Committee in charge:

Professor Jeremy Magruder, Chair
Professor Michael Anderson
Professor Max Auffhammer
Professor Stephen Raphael

Spring 2016

Abstract

Essays in Peer Effects and the Economics of Crime

by
Megan Stevenson

Doctor of Philosophy in Agricultural and Resource Economics
University of California, Berkeley
Professor Jeremy Magruder, Chair

This dissertation contains three related articles on peer effects and the economics of crime. The first article, "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes" shows how pretrial detention increases the likelihood that defendants will plead guilty and will plead to less favorable terms. The second article, titled "Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails" shows that peer effects in juvenile correctional facilities affects future crime, and discusses potential mechanisms. The third article, titled "Tests of Random Assignment to Peers in the Face of Mechanical Negative Correlation: An Evaluation of Four Techniques" presents a novel test of random assignment to peer groups and evaluates it alongside three others using simulations.

Article 1: Introductory Note

Instrumenting for detention status with the bail-setting propensities of rotating magistrates I find that pretrial detention leads to a 13% increase in the likelihood of being convicted, an effect explained by an increase in guilty pleas among defendants who otherwise would have been acquitted or had their charges dropped. On average, those detained will be liable for $128 more in court fees and will receive incarceration sentences that are almost five months longer. Effects can be seen in both misdemeanor and felony cases, across age and race, and appear particularly large for first or second time arrestees. Case types where evidence tends to be weaker also show pronounced effects: a 30% increase in pleading guilty and an additional 18 months in the incarceration sentence. While previous research has shown correlations between pretrial detention and unfavorable case outcomes, this paper is the first to use a quasi-experimental research design to show that the relationship is causal.

Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes

I have had the 'you can wait it out or take the deal and get out' conversation with way too many clients. -a public defender, Philadelphia

1 Introduction

The use of money bail has long been contentious, as it implies that freedom is reserved for those who can afford to pay for it. According to Bureau of Justice Statistics, five out of six people detained before trial are there because they could not afford their bail (Cohen and Reaves, 2007). For some, bail was intentionally set at an unaffordable level to keep them behind bars. But many have bail set at amounts that would be affordable for the middle or upper-middle class but are simply beyond the reach of the poor. In Philadelphia, the site of this study, only 51% of those who were assigned a bail amount less than or equal to $500 were able to pay the $50 deposit required for release within the first three days, and 25% remained in jail at the time of disposition. While the loss of freedom during the pretrial period is a serious concern, there are indirect consequences of pretrial detention which are potentially just as serious. If detention increases the likelihood that the defendant will plead guilty although innocent, accept an excessively-punitive plea deal, or lose at trial because detention impaired her ability to mount a successful defense, then the socio-economic disparities induced by a money bail system extend beyond the pretrial period.

Right now there is a wave of momentum in bail reform that dwarfs any seen in decades. Media attention to the recent deaths of Kalief Browder and Sandra Bland and budget pressures imposed by the 11.4 million yearly admissions to local jails have spurred jurisdictions all over the country to reconsider long-standing practices.[1][2] Yet despite all this activity,

[1] In the last several years, pretrial reform has been committed to or implemented in New Jersey, Kentucky, Colorado, Maryland, New Mexico, Connecticut, Chicago, New York City. 26 cities are implementing new pretrial risk assessment regimes in partnership with the Laura and John Arnold Foundation and 20 cities are developing pretrial reform proposals with a $75 million fund from the MacArthur Foundation.

[2] Many of the 11.4 million admissions will spend only short periods in jail. The point-in-time count of jail inmates is 750,000, two thirds of whom are awaiting trial. BJS estimates that 95% of jail growth since the year 2000 has come from pretrial detainees (Minton and Zeng, 2015).

research on the pretrial period is limited. The work that does exist is almost entirely qualitative or correlational; the few experimental or quasi-experimental studies that have been conducted either did not examine case outcomes or did not report the results in a way that allowed for causal inference (Ares et al., 1963; Goldkamp, 1980; Abrams and Rohlfs, 2011).

In this paper I present the first quasi-experimental evidence that pretrial detention increases the likelihood of being convicted, pleading guilty, being sentenced to incarceration, and being required to pay hundreds or thousands of dollars in court fees. In Philadelphia, the bail decision is made by one of six Arraignment Court Magistrates who have broad discretion in setting bail. These magistrates work on a rotating schedule which ensures that over time, the sample of defendants seen by each magistrate will be statistically near identical. By determining the bail amount, the magistrate has influence over whether a defendant is detained pretrial, but is unlikely to affect case outcomes through any other channel. All of the other activities that take place in the bail hearing are pro-forma and the schedules of the magistrates do not correlate with the schedules of the judges or attorneys that are influential later in the criminal proceedings. Thus the bail preferences of the magistrate who presides over the bail hearing provide an exogenous shock to the likelihood of being detained pretrial: a suitable instrument with which to identify the causal impact of pretrial detention on case outcomes.

The data used in this analysis covers all criminal cases originated in Philadelphia between September 2006 and February 2013. The sample size is 331,615 criminal cases with eight total bail magistrates. I use a jackknife instrumental variables technique which, in my preferred specification, allows the preferences of the magistrates to vary over time and according to offense, criminal history, race, and gender of the defendant (Mueller-Smith, 2015). This allows me to exploit considerable within-magistrate variance in detention rates. For example, a defendant with a shoplifting charge is 30 percentage points more likely to be detained if she is seen by the magistrate who is most strict on shoplifting charges instead of the one who's most lenient. However the magistrate that is most lenient on shoplifting charges is most strict on robbery charges, the magistrate that is most lenient on robbery is strict on drug possession, and so forth.

Using variation in the propensity to detain both within and across the magistrates, I estimate that pretrial detention leads to a 6.6 percentage point increase in the likelihood of being convicted on at least one charge, over a mean 50% conviction rate. The effect on conviction (being found guilty either through plea or at trial) is almost entirely explained by a 5.3 percentage point increase in the likelihood of pleading guilty among those who would otherwise have been acquitted, diverted, or had their charges dropped. Those detained will be liable for $128 more in non-bail court fees and will be sentenced to an additional 140 days of incarceration.[3]

Effects appear to be present in both misdemeanors and felonies, although the effects are much more precisely estimated in the former. Those detained on a misdemeanor charge will be 8 percentage points more likely to be convicted and 8 percentage points more likely to receive a sentence of incarceration. Effects are similar in magnitude for white and black defendants, don't vary substantially according to defendant's age, but appear largest for first or second time arrestees. For those with very limited experience in the criminal justice system, pretrial detention leads to a 17 percentage point increase in the likelihood of being convicted.

I divide the sample according to the strength of the evidence that tends to be available in different crime types. My hypothesis is that extra-legal factors such as detention will be more influential among cases where the facts are in dispute than among cases where the evidence is difficult to contest. I find that effects among 'strong-evidence' crimes (DUI, drugs, illegal firearms) are generally small and not statistically significant in the IV regressions, but effects among 'weak-evidence' crimes (assault, vandalism, burglary) are more pronounced. Those detained on a weak-evidence charge are 7 percentage points more likely to plead guilty and will receive an additional 18 months of incarceration. Since weak-evidence crimes tend to have higher rates of wrongful convictions, these findings are consistent with the claim that pretrial detention increases the likelihood that innocent people will plead guilty.[4]

Previous studies using multivariate regression techniques usually show that pretrial detention is correlated with unfavorable case outcomes even after controlling for a number of defendant and case characteristics (Ares et al., 1963; Rankin, 1964; Goldkamp, 1980; Williams, 2003; Free, 2004; Phillips, 2007, 2008; Tartaro and Sedelmaier, 2009; Sacks and Ackerman, 2012; Oleson et al., 2014; Heaton et al., 2016). Many have found very large correlations. An often-cited study by the Arnold Foundation concludes that low risk defendants who are detained throughout the pretrial period are 5.41 times more likely to be convicted and sentenced to jail than those who are released (Lowenkamp et al., 2013). They control for offense within eight broad categories as well as basic demographics and criminal history measures, but variations in charge severity, strength of the evidence, quality of the representation, wealth of the defendant, and a variety of other unobservable differences may have biased the results.

The IV estimates I present here are considerably lower than most previous estimates.

[3]Average court fees for the convicted are $775 and those who receive a prison sentence will serve almost two years on average before being eligible for parole.

[4]A forthcoming study by Charles Loeffler and Jordan Hyatt examines wrongful convictions through the use of anonymous surveys of prison inmates. Their preliminary findings suggest that the rate of wrongful convictions is higher among assaultive crimes and lower among DUIs and some drug crimes.

This may be due to omitted variable bias in previous research, differences in effect sizes across jurisdictions, or the fact that the IV estimates capture the effect of pretrial detention only for those on the margins of being detained. I also conduct OLS regressions of case outcomes on pretrial detention, controlling for narrowly defined offense categories as well as many other defendant and case characteristics. The OLS results do not differ substantially from the IV results, suggesting that researchers who are able to control for narrowly defined offense categories as well as a variety of criminal history and demographic variables may be able to produce reasonable estimates of the effect of pretrial detention on case outcomes even in the absence of a natural experiment.[5]

The impact that pretrial detention has on case outcomes could come through a variety of channels. A common claim is that people plead guilty just to get out of jail (Bibas, 2004). While this is a reasonable mechanism for misdemeanor cases, effects are also seen in much more serious charges, where it is unlikely that a defendant would be able to avoid a prison sentence. It may be that since some of the disruptions of incarceration have already occurred – loss of job/housing, the initial adjustment to life behind bars – the incentives to fight the charges are lower. Jail may affect optimism about the likelihood of winning the case, or may affect risk preferences in such a way that the certainty of a plea deal seems preferable to the gamble of a trial.[6] Detention also impairs the ability to gather exculpatory evidence, makes confidential communication with attorneys more difficult, and limits opportunities to impress the judge with gestures of remorse or improvement (taking an anger management course, entering rehab, etc.) (Goldkamp, 1980). Detained defendants are likely to attend pretrial court appearances in handcuffs and/or prison garb, creating superficial impressions of criminality. Furthermore, if a defendant must await trial behind bars he may be reluctant to employ legal strategies that involve delay. Whereas a released defendant may file continuances in the hopes that the prosecution's witnesses will fail to appear, memories will blur, or charges eventually get dropped, a detained defendant pays a much steeper price for such a strategy. More nefariously, those detained have less opportunity to coerce witnesses, destroy evidence or otherwise impede the investigation (Allen and Laudan, 2008). While pretrial detention may increase wrongful conviction, pretrial release may decrease the likelihood of successfully prosecuting the truly guilty.

While this paper does not argue for or against the use of pretrial detention, it shows

[5] In forthcoming work with Paul Heaton and Sandra Mayson (Heaton et al., 2016) we use OLS methods with extensive controls to estimate the impact pretrial detention has on case outcomes in Harris County, Texas. In general, results are similar to those found in Philadelphia; effect sizes are slightly larger but this may be explained by institutional differences across the different jurisdictions.

[6] Detention may change the reference point so that freedom is seen as a gain instead of incarceration being seen as a loss. Prospect theory would thus predict a shift towards more risk averse decisions (Daniel Kahneman, 1979).

that detention has significant downstream consequences. Being found guilty in the court of law comes with hefty court fees, periods of incarceration or probation, potential disenfranchisement, a criminal record, and challenges securing future employment, school loans, or housing (Uggen et al., 2012; SHRM, 2012; Thacher, 2008; ONDC, n.d.). If wealth or race influence the likelihood of being detained pretrial – and both previous research as well as evidence presented in this paper suggest that they do – then pretrial detention exacerbates socio-economic inequalities in the criminal justice system (Schlesinger, 2005; Wooldredge, 2012).

In Section 2 I give a brief overview of the pretrial process, in Section 3 I describe the natural experiment, and in Section 4 I discuss the data and provide descriptive statistics and graphs. Section 5 tests the identifying assumption and Section 6 discusses the empirical strategy and shows a visual representation of the first and second stage. Section 7 presents the results for the full sample and Section 8 shows results for various subgroups. Section 9 provides robustness tests and shows how results vary with the number of days detained. Section 10 concludes.

2 Bail, pretrial detention, and an overview of criminal proceedings

Pretrial detention is the act of keeping a defendant confined during the period between arrest and disposition for the purposes of ensuring their appearance in court and/or preventing them from committing another crime. The vast majority of jurisdictions use a money bail system to govern whether or not a defendant is detained. In such a system a judge or a magistrate determines the amount of the bail required for release and the defendant is only released if she pays that amount and agrees to certain behavioral conditions. In some cases the defendant will be released without having to pay anything, in others (usually only the most serious cases) she will be denied bail and must remain detained. While the defendant is liable for the full amount of the bail bond if she fails to appear at court or commits another crime during the pretrial period, she usually does not need to pay the full amount in order to secure release. In many jurisdictions she will borrow this sum from a bail bondsman, who charges a fee and holds cash or valuables as collateral. In some jurisdictions, Philadelphia included, the courts act as a bail bondsman and will release the defendant after the payment of a deposit.

Bail hearings are generally quite brief – in Philadelphia most last only a minute or two – and often do not have any lawyers present. After the bail hearing there are a series of pretrial court appearances that defendants must attend. Although the exact procedure varies across jurisdictions these usually include at least an arraignment (where formal charges are filed) and some sort of preliminary hearing or pretrial conference (where the case is discussed and plea deals can be negotiated). Plea bargaining usually begins around the time of arraignment and can continue throughout the criminal proceedings. In some jurisdictions, like New York City, the arraignment happens simultaneous to the bail hearing and it is not uncommon to strike a plea deal at this first appearance. In other jurisdictions, such as New Orleans, arraignments often do not happen until six months after the bail hearing and a defendant who is unable to make bail must wait until then to file a plea. In Philadelphia, arraignments usually happen within a month of the bail hearing.

Plea negotiation is a process in which the defendant receives reduced charges or shorter sentences in return for pleading guilty and waiving her right to a trial. Since defendants often face severe sentences if found guilty at trial, the incentives to plead are strong. It's estimated that 90-95% of felony convictions are reached through a plea deal (Devers, 2011). Philadelphia differs from many other jurisdictions in its wide use of bench trials on felony cases. Since sentencing tends to be more lenient in bench trials than jury trials, this reduces the incentive to plead guilty.[7] As such, only about 75% of felony convictions are reached through plea in Philadelphia. Trial by jury is not constitutionally required if the maximum incarceration sentence is less than six months, and the use of bench trials for misdemeanors, as is the custom in Philadelphia, is more common across jurisdictions.

This paper shows that pretrial detention increases the likelihood that a defendant will plead guilty and will plead to less favorable terms. While there is little reason to believe this result is unique to Philadelphia, the magnitude of the effect is likely to differ across jurisdictions due to variations in the criminal justice proceedings.

3 The natural experiment

Immediately after arrest, arrestees are brought to one of seven police stations around the city. There, the arrestee will be interviewed via videoconference by Pretrial Services. Pretrial Services collects information about various risk factors as well as financial information to determine eligibility for public defense. Using risk factors and the current charge, Pretrial Services will determine the arrestee's place in a 4 by 10 grid of bail recommendations. These bail guidelines suggest a range of appropriate bail, but are only followed about 50% of

[7]In Philadelphia, a bench trial is the default for all but the most serious felonies. The right to a jury trial can be asserted upon request, but this is uncommon. While there is no formal mechanism that ensures that a bench trial will lead to better outcomes for the defendant than a jury trial, all defense attorneys interviewed assured me that this was the case.

the time (Shubik-Richards and Stemen, 2010). Once Pretrial Services has entered the bail recommendation and the financial information into the arrest report the arrestee is ready for her bail hearing.

Once every four hours the magistrate will hold bail hearings for all arrestees on the 'list' (those who are ready to be seen). The bail hearing will be conducted over videoconference by the magistrate, with a representative from the district attorney's office, a representative from the Philadelphia Defender Association (the local public defender), and a clerk also present. In general, none are attorneys. The magistrate makes the bail determination on the basis of information in the arrest report, the pretrial interview, criminal history, bail guidelines, and advocacy from the district attorney and public defender representatives.

There are four things that happen during the bail hearing: the magistrate will read the charges to the arrestee, inform her of her next court appearance, determine whether the arrestee will be granted a court-appointed defense attorney, and set the bail amount. The first two activities are formalities that ensure the defendant is aware of what she is being charged with and where her next court date is. Eligibility for public defense is determined by income. If the defendant is deemed eligible, she will be assigned either to the Defender Association, or to a private attorney who has been approved to accept court appointments by the City of Philadelphia. The default is to appoint the Defender Association; if procedural rules require the court to appoint an attorney outside of the Defender Association the magistrate's clerk will appoint the attorney at the top of a rotating list of eligible attorneys known as a 'wheel'.[8]

A typical bail hearing lasts only a minute or two and the magistrate has broad authority to set bail as she sees fit.[9] Bail decisions fall into three categories: release with no payment required, cash bail or no bail.[10] Those with cash bail will be required to pay a 10% deposit on the total bail amount in order to be released. After disposition, and assuming that the behavioral conditions of the pretrial period were met, 70% of this deposit will be returned. The City of Philadelphia retains 30% of the deposit, even if charges get dropped or the defendant is acquitted on all charges. Those who do not have the 10% deposit in cash can

[8]If there are multiple codefendants, such that representing all of them would pose a conflict of interest, one defendant will be randomly selected to be served by the Defender Association and the others will receive a court-appointed attorney. For opaque historical reasons, four out of five defendants charged with murder will be represented by court-appointed attorneys and the fifth will be represented by the homicide division of the Defender Association (Anderson and Heaton, 2012). This decision is made by the order in which defendants are entered into the data system and the court-appointed attorney is chosen by a Municipal Court judge, not a magistrate.

[9]If either the defense or the prosecution is unhappy with the decision they can make an appeal to a judge immediately after the bail hearing. However the bar is high for overturning the original bail decision so this is not very common.

[10]Holding a defendant without bail is uncommon, although bail is sometimes set at prohibitively high rates.

borrow this amount from a commercial bail bondsman, who will accept cars, houses, jewelry and other forms of collateral for their loan. If the defendant's arrest occurred while she is already on parole, her parole officer may choose to file a detainer. If a detainer is filed she may not bail out until a judge removes the detainer.[11]

The research design uses variation in the propensity of the magistrates to assign affordable bail as an instrument for detention status. The validity of the instrument rests on several factors, including that the magistrate received is essentially random and that the instrument not affect outcomes through a channel other than pretrial detention. The following details help ease concerns along these lines.

Philadelphia employs six Arraignment Court Magistrates at a time, and one of the six will be on duty 24 hours a day, 7 days a week, including holidays. Each day is composed of three work shifts: graveyard (11:30 pm-7:30 am), morning (7:30 am-3:30 pm) and evening (3:30 pm-11:30 pm). Each magistrate will work for five days on a particular shift, take five days off, then do five days on the next shift, five days off, and so forth. For example, a magistrate may work the graveyard shift from January 1st to January 5th, have January 6th-10th off, then work the morning shift from January 11th-15th, have the 16th-20th off, do the evening shift from January 21st-25th, take the next five days off, and then start the cycle all over again.

This rotation relieves concerns that certain magistrates set higher bail because they work during shifts which see higher-risk defendants. Over time, each magistrate will be scheduled to work a balanced number of weekends, graveyard shifts, and so forth. However the magistrates do not always work their appointed shifts; in fact, about 20% of the time there is a substitute (usually one of the other magistrates). To avoid potential confounds I instrument with the magistrate who was scheduled to work instead of the magistrate that actually worked. Furthermore, arrestees do not have latitude to strategically postpone their bail hearing to receive a more lenient magistrate. The process from arrest to bail hearing has been described as a conveyor belt: on average the time from arrest to the bail hearing is 17 hours and defendants are seen as soon as Pretrial Services notifies the Arraignment Court that they are ready (Clark et al., 2011). Thus the cases seen by each scheduled magistrate should be statistically very close to identical. I confirm this empirically in Section 6.

Since the duties of the bail magistrate are so limited, there are few channels outside of the setting of bail through which the magistrate could affect outcomes. One concern would be a correlation between the schedules of the magistrates and the likelihood of receiving a particular judge, prosecutor or defense attorney later on in the criminal proceedings. However

[11] The detainer hearing usually happens within a week of arrest. Detainer cases are evenly distributed across magistrates and should not bias the results.

the peculiar schedule of the magistrates does not align with the schedule of any other actors in the criminal justice system. For one, this is because the other courts are not open on weekends. This is also because Philadelphia predominantly operates on a horizontal system, meaning that a different prosecutor handles each different stage of the criminal proceedings. Likewise, if the defendant is represented by the Defender Association (∼60% of the sample), she will have a different defense attorney at each stage.[12] While attorneys often rotate duties, their rotations are based on a Monday-Friday work week and not the 'five days on, five days off' schedule of the magistrates.

While eligibility for public defense is another potential channel through which the magistrate could affect outcomes, this is supposed to be a pro-forma action based on the income of the defendant. I am not able to see whether the defendant is deemed eligible, I only know the attorney type at the time of disposition. This is likely to be affected by detention status, as the decision to hire a private attorney is based both on having the money to pay one and expectations over case outcomes. While I cannot test directly whether magistrates vary in the rates at which they grant eligibility, I conduct robustness checks in which I control for the attorney type at the time of disposition. This has only a trivial impact on the results.

4 Data and descriptive statistics

The data for this analysis comes from the court records of the Pennsylvania Unified Judicial System; PDF files of case dockets and criminal histories are publicly available online. The data covers all criminal cases which had a bail hearing between September 13, 2006 and February 18, 2013.[13] Before September 13, 2006, Philadelphia used a different data management system with much lower data quality. I do not look at cases which began after February 18, 2013 both because I wanted to leave ample time for all cases to resolve and because one of the magistrates was replaced by a new one on that date.

Figure 1a shows a histogram of the number of days defendants are detained before disposition, conditional on being detained more than three days and less than 600 days.[14] The left tail of the distribution is omitted since the primary definition of 'detainees' used in this paper is being unable to make bail within three days; the long right hand tail of the distribution is omitted for visual simplicity. The median number of days detained for those who are unable to make bail within three days is 78, the mean is 146.

[12]The most serious cases are not handled horizontally, however the choice of attorney to handle these cases has nothing to do with the magistrate.

[13]Case outcome or bail data was missing for 0.44% of the sample. These observations were dropped.

[14]Although the sixth amendment guarantees the right to a speedy trial this can be waived if the defendant finds it beneficial to extend.

Summary statistics for the released group, the detained group, and the whole sample are shown in Table 1. Defendants are predominantly male, and, although race is missing for 11% of the sample, largely African-American. Those detained tend to have longer criminal histories and are facing more serious charges than those released. It should be noted, however, that 25% of the detained sample are only facing misdemeanor charges.[15]

Almost half the sample have their charges dropped, dismissed, or are placed in some sort of diversion program.[16] Almost everyone else was convicted, through plea or at trial, on at least one charge. 90% of cases resolved at trial result in convictions, suggesting that prosecutors will not bring a case to trial if they don't believe they have a strong chance of winning. If a detained defendant pleads quickly to avoid more time waiting in jail, she may be pleading guilty on a case that otherwise would not have proceeded to court.

One third of the sample is released without being required to pay bail and an additional 26% are able to pay their way out within three days of the bail hearing. Figure 1b shows the number of people detained and released at a variety of bail ranges. The median amount of bail for the detained group is $10,000. Almost 40% of those who were given bail amounts less than or equal to $2000 are detained for longer than three days. Among this low-bail sample – 77% of whom are charged only with misdemeanors – the average number of days detained pretrial is 28. This group would need to pay a deposit $200 or less to secure their freedom.

Table 2 shows the most common lead charges. The table is organized in descending order, where the top of the list represents crime types where evidence tends to be strongest. The horizontal line separates those classified as 'strong-evidence' crimes from those classified as 'weak-evidence' crimes. The ranking of charges by strength of evidence was done in two ways. First, I surveyed a variety of Philadelphia lawyers and professors who specialize in criminal law. I asked them to rank each crime on a scale of one to five, where a one meant that the evidence in that crime type tended to be ambiguous and subject to multiple interpretations and a five meant that the evidence in that crime type tended to be clear and difficult to dispute. Second, I calculated the average conviction rate per crime type under the assumption that the conviction rates would be higher among crime types where the evidence was stronger. I ranked each crime according to both measures of evidence strength; the order in which they are presented here is the average of the two rankings. While there were

[15]The offense information used in this paper is taken from the charge at the time of the bail hearing. Many of those who were originally charged with felonies subsequently had the felony charge downgraded to a misdemeanor.

[16]Diversion programs are designed for those with low level misdemeanor charges; if the defendant agrees to requirements such as paying restitution to victims, entering rehab, or performing community service, they are generally able to avoid a formal adjudication of guilt and a criminal record.

differences across the two methods in the exact ordering, the general placement was quite similar.[17]

Figures 1c shows the likelihood of pleading guilty at various levels of 'sentence exposure'. Sentence exposure is a measure of how serious the case is: the average incarceration sentence that similarly situated defendants receive if they are found guilty at trial. This is estimated by taking the fitted values from a regression of sentence length on offense, criminal history, demographics, and time controls, using the sample of cases in which the defendant was found guilty at trial. A log transformation is applied to the fitted values to compress the long right tail. Figure 1c shows that while the likelihood of pleading guilty is relatively flat at low levels of sentence exposure, it rises rapidly and then falls off at higher levels of sentence exposure. While no definitive conclusions should be inferred from this descriptive graph, it is consistent with a story in which defendants plead guilty out of fear of a sentence exposure at trial. While the plea rate drops off at the highest levels of sentence exposure, this may be because the guilty pleas themselves come with increasingly long prison sentences. A present-biased defendant may underweight the difference between the 10 year sentence she is offered via plea or the 25 year sentence she is at risk of receiving at trial.

Figure 1d shows conviction rates at various levels of sentence exposure. It shows that throughout most of the support of the distribution, the likelihood of being convicted decreases with sentence exposure. Figure 1e shows the likelihood of detention at various levels of sentence exposure. We see a monotonically increasing relationship between sentence exposure and the likelihood of being detained. This is not surprising, as those who face longer sentences will face increased incentive to flee, or may pose greater public safety risks to the community.[18]

In Table 3 I test to see if there are socio-economic disparities in the likelihood of being detained pretrial. I regress the log bail amount and a dummy for pretrial detention on race and the log of average income in the defendant's zip code, controlling for offense, criminal history, age and gender.[19] I limit the sample to those for whom both zip code and race is available.[20] I find that those coming from wealthier neighborhoods have bail set lower and are less likely to be detained pretrial even after controlling for a wide range of characteristics. A

[17]The ranking based on the survey would have placed car theft in the 'strong-evidence' category and other types of theft in the 'weak-evidence' category.

[18]These three graphs, taken together, suggest that OLS regressions of case outcomes on detention status may in fact be biased downwards. Since the seriousness of the case is not fully visible in the data (offense can be controlled for, but severity can vary within charge categories) this omitted variable will increase the likelihood of detention but may actually decrease the likelihood of conviction. Other likely omitted variables, such as wealth, lawyer quality, and community ties are more consistent with an upward bias of OLS estimates.

[19]I add one to all bail amounts to avoid taking the log of zero.

[20]Average gross income per zip code in 2010 was downloaded from IRS.gov.

10% increase in zip code wealth is correlated with a .4% decrease in bail and a .2 percentage point decrease in the likelihood of being detained pretrial. Race does not correlate with the bail amount in a statistically significant manner, however African-Americans are three percentage points more likely to be detained pretrial than Caucasian defendants with similar offense profile, age, gender, and criminal history.

5 Identification test

Given the rotation of the work schedules, there may be slightly more imbalance across the magistrates than there would be if the magistrate was randomly and independently drawn for each bail hearing. Since most parametric tests of randomness assume an independent random draw, I conduct a permutation test to verify that the sample of defendants seen by each magistrate are no more different from one another than would be expected by chance, given a rotating work schedule such as the one in use.

I generate a variety of "false" work schedules, keeping the basic parameters of the work rotation fixed: five days on, five days off, rotating shifts, three eight hour shifts per day, etc. However I vary both the day at which the five day rotation begins, the hour at which the shifts begin, and the direction of rotation. (The actual magistrates move from graveyard to morning to evening shift, a reverse rotation would move from graveyard to evening to morning.) Since the schedule is five days long there are four potential "false" start dates to choose from. Since the shift is eight hours long there are three potential "false" start times, spaced two hours apart. I generate three false rotations: one in which both groups work a reverse rotation, and two in which one of the groups work a forward rotation while the other group works a reverse rotation. Given five start dates (four false and one real), four start times (three false and one real) and the three false rotations (I do not use the real rotation to minimize correlations with the actual work schedule) I build 60 false work schedules.

$$Cov_i = \alpha + \beta * Magistrate_i + \psi * Time_i + e_i \tag{1}$$

With each of these work schedules I run the regression specified in Equation 1, where Cov_i is one of 66 different offense, criminal history and demographic covariates, $Magistrate_i$ is the dummy for the magistrate who was scheduled during the bail hearing for case i under the false schedule, and $Time_i$ are the full set of controls for the time and date of the bail hearing as described in Section 4.[21] For each covariate and each false work schedule I collect

[21]Unless otherwise specified, all regressions shown in this paper control for the variables listed as follows. Demographic variables include age, age squared, age cubed, gender, and dummies for being black or white. Criminal history variables include the number of prior cases, prior felony cases, prior cases involving a serious

the F statistic for the joint magistrate dummies. This creates an empirical distribution of the F statistic under false schedules. I then compare the F statistic from the Equation 1 regression in which the magistrate dummies are those who were *actually* scheduled to the distribution of F statistics under false schedules. For each covariate I build an 'empirical p value' which is the fraction of false-schedule F statistics that are larger than the true-schedule F statistic.

I also perform this technique on three summary variables which are the fitted values from a regression of a dummy for pretrial detention, a dummy for pleading guilty to at least one charge, and a dummy for being convicted on at least one charge on all 66 of the offense, criminal history, demographic and time controls. These predicted values are a weighted average of the characteristics that correlate most strongly with the main dependent variables and the endogenous independent variable. I show the F statistic and the empirical p value for these three summary statistics in Table 4. The F statistics are 1.84, 2.59 and 1.91 respectively; as expected the covariates are somewhat less balanced across the magistrates than they would be if the bail magistrate was independently and randomly assigned to each bail hearing. However, at 0.96, 0.56 and 0.58 respectively, the empirical p values show that any imbalance in the predicted likelihood of detention, guilt, or guilty pleas across the magistrates is no more than would be expected by chance.

violent crime (robbery, assault, burglary, murder, rape), prior cases where the defendant was found guilty of at least one charge, and dummies for having at least one prior case, having at least three prior cases, awaiting trial on another charge, having a detainer placed on them, and having a prior arrest within five years of the bail hearing. Offense variables include dummies for having a charge in the following category: murder, rape, robbery, aggravated assault, burglary, theft, shoplifting, simple assault, drug possession, drug selling, drug buying, possession of marijuana, F2 firearm, F3 firearm, possession of stolen property, vandalism, a non-firearm weapon charge, prostitution, first offense DUI, second offense DUI, resisting arrest, stalking, motor vehicle theft, indecent assault, arson, solicitation of prostitutes, disorderly conduct, pedophilia, intimidation of witnesses, accident due to negligence, false reports to a police officer, fleeing an officer, and reckless endangerment. Additional offense controls include dummies for being charged with a first, second or third degree felony, an unclassified felony, a first, second or third degree misdemeanor, an unclassified misdemeanor, or a summary offense. I also control for the total number of charges, the total number of felony charges, the total number of misdemeanor charges, and the total 'offense gravity score' of the charges (the offense gravity score is used by Philadelphia to measure the seriousness of a charge on a scale of 1-8). Time controls include dummies for the day of the week that the bail hearing occurs, a dummy for graveyard, morning, and evening shift, a cubic in day of the year (1-365), the bail date, and year dummies. Since magistrates tend to leave and be replaced in the third week of February, I define the year dummies to align with the start and end of a magistrate's work period. The following dates serve as dividers: February 21, 2008, February 23, 2009, February 23, 2010, February 23, 2011, February 23, 2012, and February 18, 2013. I interact some of the covariates with three time periods, as divided by February 23, 2009 and February 23, 2011. The covariates that I allow to have differing impacts over these three time periods are the same as the ones that I interact with the magistrate fixed effects in various specifications: murder, robbery, aggravated assault, burglary, theft, shoplifting, simple assault, drug possession, drug sale, drug purchase, marijuana possession, F2 firearm, F3 firearm, vandalism, prostitution, first offense DUI, motor vehicle theft, gender, a dummy for being African-American, the number of prior cases, the number of prior violent crimes, a dummy for having at least one prior and a dummy for having a detainer.

Figure 1f shows a histogram of the 69 empirical p values from the 66 covariates and three summary statistics. As can be seen, the empirical p values are evenly spread between zero and one. The mean p value is 0.56, the median is 0.58, and three p values are less than or equal to 0.05. The mean F statistic is 2.21 and the median F statistic is 1.84.

The 'false' F statistics comprising the empirical distribution will be correlated, since there is overlap in the false schedules. This reduces the power on any single test (any single covariate) since the 60 different false F statistics are not independent from one another. However among the 69 different tests there is no evidence to suggest strategic behavior that would undermine the credibility of the research design.

6 Empirical strategy and visual representations of the research design

Instrumenting for sentencing outcomes using varying propensities of randomly assigned or rotating judges is a popular method of identifying the impact that the criminal justice system has on defendants (Kling, 2006; Aizer and Doyle, 2009; Loeffler, 2013; Tella and Schargrodsky, 2013; Mueller-Smith, 2015). Two concerns with this methodology are that the judges may affect outcomes through channels other than the primary sentencing characteristics the researcher is interested in, and that the monotonicity assumption may be violated. Concern about alternate channels of influence are minimized in this setting due to the limited nature of the bail magistrates' responsibilities and the fact that their work schedule does not align with the other actors in the criminal justice system. However the assumption that the bail-setting propensities for each magistrate do not vary according to defendant characteristics would be harder to defend.

I follow Mueller-Smith (2015) in allowing the magistrates' bail habits to vary across time and according to defendant and case characteristics. Allowing the preferences of the magistrates to update every two years relaxes the assumption that all magistrates respond in the same way to changes in the criminal justice system: a new district attorney, an update to the bail guidelines, a change in capacity constraints at the local jail. Allowing the preferences of the magistrates to vary across case and defendant characteristics relaxes the assumption that a strict magistrate must be equally strict on all crimes and all defendants. In addition to minimizing non-monotonicity bias, exploiting within-magistrate variation in bail-setting propensities increases power.

Empirically, allowing the magistrates' bail preferences to vary is accomplished by interacting the magistrate fixed effects with time period fixed effects and a subset of the covariates in the first stage regression, as shown in Equation 2. $Detention_i$ is a dummy which is equal to one if the defendant is detained for more than three days in case i, $Magistrate_i$ is a dummy for the magistrate who was scheduled to work during the bail hearing for case i, and Cov_i^1 are the subset of the covariates across which I allow the magistrate's preferences to vary: the 16 most common crime types, gender, race, and criminal history. Cov_i^2 are the remainder of the offense, demographic and criminal history controls, as listed in Footnote 21. T_i^3 is a dummy for the time period of the bail hearing for case i (there are three time periods as divided by February 23, 2009 and February 23, 2011) and $Time_i$ are the full set of controls for the time and date of the bail hearing (which include T_i^3 as a subset). I use a jackknife instrumental variables technique to avoid the bias induced by many instruments (Angrist et al., 1999).

$$\begin{aligned} Detention_i = \alpha_2 + Magistrate_i * T_i^3 * \omega_2 + Magistrate_i * Cov_i^1 \\ * \phi_2 + Cov_i^1 * T_i^3 * \delta_2 + Cov_i^2 * \gamma_2 + Time_i * \psi_2 + e_i \end{aligned} \quad (2)$$

The second stage is shown in Equation 3 where $Case\_Outcome_i$ represents a variety of case outcomes, $\widehat{Detention}_i$ is the fitted value from the jackknifed first stage, and Cov_i^1, Cov_i^2, T_i^3 and $Time_i$ are as described above. I present results from a linear regression in both stages; logit regressions yield similar results.

$$Case\_Outcome_i = \alpha_3 + \widehat{Detention}_i * \beta_3 + Cov_i^1 * T_i^3 * \delta_3 + Cov_i^2 * \gamma_3 + Time_i * \psi_3 + \epsilon_i \quad (3)$$

Figures 2 and 3 show visual representations of the first and second stage of the IV regression. Figure 2 shows how pretrial detention rates vary by magistrate for different crime types. The y axes show residuals from a regression of pretrial detention on controls for the time and date of the bail hearing. The eight bars represent the average detention residuals for the eight magistrates, ordered so that Magistrate 1 has the lowest overall rate of pretrial detention, Magistrate 2 has the second lowest, and so forth. The order of the magistrates remains the same in all subplots. Figure 2a shows the magistrate means across the entire sample, Figures 2b-f show the means for different crime types. In each offense category, n refers to the number of cases charged with that offense and F refers to the joint significance of the magistrates in a regression of detention status on magistrate dummies and time controls on that sub-sample. The figures demonstrate that while magistrates vary in their overall detention rates, there is quite a bit of variance across the different crime types. For example, although the overall detention rate of Magistrates 2 is quite low, this magistrate has the highest detention rates for prostitution and drug possession. This within-magistrate variation in detention rates suggest that the monotonicity assumption will be violated if detention status is instrumented for with the overall detention rate of the magistrates.

Figure 3 shows that defendants whose bail hearing is presided over by high-bail-setting magistrates are more likely to be convicted or to plead guilty. In Figure 3a the y and x axes show residuals from a regression of conviction and detention dummies respectively on the set of time controls described by $Time$. Figure 3b is similar except the dummies are residualized over $Cov^1 * T^3, Cov^2$ and $Time$. Each circle represents the average detention and conviction residuals of the eight magistrates; the size of the circle is proportional to the total number of cases seen by each magistrate. As can be seen there is a clear positive correlation between the conviction and detention residuals which changes very little once the effect of covariates have been removed. This provides visual evidence that the slight differences in the groups of defendants seen by each magistrate are not the cause of the positive relationship between detention and conviction.

Figures 3c and d show the relationship between pretrial detention and pleading guilty for those who are charged with weak-evidence crimes and those charged with strong-evidence crimes respectively. In this graphic, the magistrate dummies have been interacted with T^3, dummies for the three time periods. The x and y axes in both figures show residuals from a regression of pretrial detention and guilty pleas respectively on covariates, time controls, and time-covariate interactions. The circles show the average residuals per-magistrate-per-time period. A clear positive relationship can be seen between the likelihood of being detained and the likelihood of pleading guilty in weak-evidence cases. There is no visually discernible relationship among strong-evidence cases.

7 Full sample results

In Table 5 I show results from a variety of jackknife instrumental variables specifications where the outcome variable is a dummy which equals one if the defendant is convicted on at least one charge in Panel A, and a dummy variable which equals one if the defendant pled guilty to at least one charge in Panel B. In Column 1 the only instruments are the eight magistrate dummies and in Column 2 the magistrate dummies are interacted with T^3, the three time period dummies. The only controls in the first two columns are $Time$. The standard errors decrease between the first and second column by about 10%, suggesting that allowing the magistrates to respond differently to the various changes that occur during the period of my analysis increases the power of the instrument. Covariates are added in Column 3 and the effect sizes either increase (Panel A) or remain constant (Panel B). Columns 4-6 allow the bail setting habits of the magistrates to vary according to offense, criminal history and demographics of the defendant. In Column 4, the magistrate dummies are interacted with dummies for the five most common lead charges: drug possession, first offense DUI, robbery, selling drugs, and aggravated assault. In Column 5 I add interactions between magistrate dummies and the number of prior cases/prior violent crimes, dummies for having at least one prior case, having a detainer, and being African-American or female. In Column 6 I add interactions between the magistrate dummies and the other most common crime types as listed in Table 2. The number of instruments per specification as well as the F statistic of joint significance on the first stage instruments are included in the bottom panel.

Both effect sizes and standard errors decrease as instruments are added. This suggests two things: that allowing the bail-setting habits of the magistrates to vary across defendant characteristics both increases the power and reduces non-monotonicity bias in the results. In particular, if treatment effects are smaller among crime types for which the monotonicity assumption is violated, then the estimates in Columns 1-3 will be biased upwards. It should be noted, however, that non-monotonicity bias will not generate spurious results if no treatment effects exist. Under the null hypothesis it would be very unlikely to see effect sizes as large as those shown in Table 5.

My preferred specification is the one where magistrate's preferences are allowed to vary across all 16 of the most common crime types, across the criminal history, race, and gender of the defendant, and over the three time periods. I estimate that pretrial detention leads to a 6.6 percentage point increase in the likelihood of being convicted and a 5.3 percentage point increase in the likelihood of pleading guilty. Compared to the means for each dependent variable, that converts into a 13% increase in the probability of conviction and a 21% increase in the likelihood of pleading guilty.

Table 6 shows how pretrial detention affects conviction rates, guilty pleas, court fees, the likelihood of being incarcerated, and both the maximum and minimum incarceration sentence.[22] Panel A shows results from the jackknife instrumental variables method with the most fully interacted specification and Panel B shows results from an OLS regression controlling for the full set of offense, criminal history, demographic and time controls. With the exception of court fees and the incarceration dummy, results do not vary substantially between IV and OLS. This suggests that researchers who are interested in estimating the effects of pretrial detention in other jurisdictions may be able to achieve reasonable results with standard court data even in the absence of a natural experiment.

The IV estimates suggest that pretrial detention leads to an average increase of $128 in non-bail court fees owed. Conditional on being convicted, court fees average at $775, and $1250 if convicted of a felony. For the tens of thousands of people convicted as a result

[22]Sentence length is coded as zero for individuals who do not receive an incarceration sentence.

of pretrial detention – many of whom were unable to pay even fairly small amounts of bail – these court fees may pose a significant burden. The IV results for the likelihood of being incarcerated and the maximum incarceration sentence are positive but noisy, however the estimates on the minimum incarceration sentence are more precise. Pretrial detention leads to an expected increase of 140 days in sentence length before being eligible for parole. Conditional on receiving a sentence of incarceration, defendants spend an average of 22 months in jail before being eligible for parole.

In results not shown here, I test to see if pretrial detention affects post-disposition crime. The estimates are noisy but are generally negative, suggesting that some crimes may be averted as a result of increased incapacitation. Forthcoming work will provide a cost-benefit analysis of the impact pretrial detention has on pretrial crime, including novel estimates of the cost of incarceration to the incarcerated (Mayson and Stevenson, 2016).

8 Results for various subsamples

In Table 7 I show results for misdemeanors and felonies using both IV and OLS techniques; the discussion below focuses on the IV results.[23][24] The effect sizes of the felony sample are similar in magnitude to the full sample, but are noisy. The effects among misdemeanors are more precisely measured and are slightly larger than the full sample estimates, especially in relation to the means of the dependent variables. In fact, pretrial detention among misdemeanor defendants leads to a statistically significant increase in all outcomes. The effects on punishment are particularly large: those detained will be 8 percentage points more likely to receive a sentence of incarceration over a mean 16 percentage point incarceration rate. While the expected increase in sentence length is only a month or two, this represents more than a 100% increase relative to the mean.[25] Those who are given an incarceration sentence for a misdemeanor crime will spend an average of 9 months in jail before being eligible for parole.

In Table 8 I compare effect sizes across weak-evidence crimes and strong-evidence crimes. Many of the strong-evidence crimes are possession crimes, where drugs or illegal firearms were

[23]The felony sample is defined as those who were charged with at least one felony at the time of the bail hearing; many of these had their charges downgraded to misdemeanors only by the time of the arraignment.

[24]The IV specifications allow the magistrate preferences to vary across time and across defendant characteristics, as shown in Column 6 of Table 5.

[25]This increase in sentence length could not be explained solely by detained defendants being released with time-served. If all of the defendants who were convicted as a result of detention were given time served this would result in an average increase of only 12.24 days. (The average number of days detained for those unable to make bail within 3 days is 144, I estimated that detention leads to a 8.5 percentage point increase in conviction. $144 * 0.085 = 12.24$)

found on the defendant's body or in her home or car. Shoplifting crimes usually entail store merchandise found on the defendant's person as they walk out of the store. In DUI's, the defendant was found behind the wheel with an elevated blood alcohol level, and prostitutes are usually arrested after soliciting from an undercover officer. The main piece of evidence among strong-evidence crimes is the police officer's statement and/or a lab report; these types of charges can be difficult to contest.

Most weak-evidence crimes are complainant offenses (an arrest was made as a result of a complaint). Evidence in these types of crimes are much more likely to consist of victim testimony, eyewitness accounts, surveillance/bystander video, alibis, character testimonials, identification from police lineups, etc. With this type of evidence it can be harder to discern the facts of the case. Assaultive crimes in particular may have multiple conflicting accounts of what occurred and why. Witness testimony can be inconsistent, videos blurry, alibis hard to verify, and police lineup identification is notoriously unreliable. While a number of theft crimes are included in this category, possession of a stolen object does not automatically imply culpability. For example, passengers in a stolen car may not be aware that the car was stolen.

My hypothesis is that extra-legal factors such as detention status will have less effect on cases where the evidence is strong than they will in cases where the facts are difficult to discern. Prosecutors are unlikely to drop charges if the evidence is strong, nor will they lose if the case is brought to trial. While a detained defendant may plead guilty sooner or to more unfavorable terms, the effect on conviction should be minimal. In a weak-evidence case, however, the defendant's willingness to wait may prove important. Cases where the evidence is weak are much more likely to be dropped, or to result in acquittal at trial. Furthermore, such cases may rely on the testimony of witnesses who are reluctant to cooperate or whose memory fades over time. In fact, if the prosecution's key witnesses fail to appear four times in a row, the case is dropped. If detention leads defendants to plead guilty or move to trial quicker than they would otherwise this could be a significant disadvantage.

Detention status may also directly affect the evidence available among weak-evidence crimes. 75% of the sample is represented by a public or court appointed attorney; the high case volumes handled by these attorneys suggest that they may not have the time to do as much investigative work as is necessary. A released defendant can contact eyewitnesses, secure surveillance video, take photos of the crime scene, and otherwise collect exculpatory evidence. A released defendant can also pressure witnesses, destroy evidence, or otherwise impede the investigation.

While the standard errors in the IV estimates are large enough that definitive conclusions can't be drawn, the results generally suggest that effect sizes are larger among weak-evidence crimes. The difference is particularly striking for sentence length: those detained on a weak-evidence crime can expect to be sentenced to an additional 18 months in prison before being eligible for parole. With the exception of court fees, the IV effects for strong-evidence crimes are close to zero and are not statistically significant.

The OLS results also support the hypothesis that effect sizes are larger among weak-evidence crimes. The OLS estimates of the effect pretrial detention has on conviction are 0.098 among weak-evidence crimes and 0.007 among strong-evidence crimes. Figure 4 shows OLS effects by crime category (the IV results are too noisy for such small samples). Strong-evidence crimes are at the top and weak-evidence crimes are at the bottom. The coefficient plot, which shows the estimated effect that pretrial detention has on guilty pleas, again suggests that effect sizes are larger among weak-evidence crimes.

Table 9 shows IV results for blacks, whites, young defendants, older defendants, those with one or no prior arrests, and those with more extensive criminal history. Overall, there is nothing to suggest that effect sizes differ across race. The point estimates are generally quite similar, although the subsample is small enough that many are not statistically significant. Results are also similar among younger and older defendants. The point estimates for sentence outcomes are greater among the younger defendants, but the standard errors are large as well. We do, however, see suggestive evidence that effect sizes are larger for those with limited prior interactions with the criminal justice system. Among first or second time arrestees, pretrial detention leads to a 12 percentage point increase the likelihood of pleading guilty and a 17 percentage point increase in the likelihood of being convicted.[26]

9 Robustness checks and effect sizes for varying definitions of pretrial detention

In Table 10 I present several robustness checks for the full sample results. Panel A is identical to Panel A of Table 6 except that magistrate dummies are also included as controls in the second stage regression. Controlling for the eight magistrate dummies implies that the impact pretrial detention has on case outcomes is being identified solely off of within-magistrate variations in detention rates. In particular, these controls will absorb any other fixed aspects of the magistrates that may affect the results. For example, the most lenient magistrate may also be particularly encouraging or supportive during the bail hearing. If this affects the defendant's expectations of success at trial – and thus their willingness to accept a guilty plea – this would undermine the exogeneity assumption. If

[26]The sample of first time arrestees is so small that the IV results are hard to interpret.

the effect sizes change greatly as a result of including magistrate fixed effects as controls this would raise concerns that the magistrates are affecting case outcomes through channels other than pretrial detention. Panel A shows that although the inclusion of magistrate fixed effects increases the standard errors, the effect sizes are not changed dramatically.

Panel B is identical to Panel A of Table 6 except that controls for attorney type are added. While attorney type is likely to be endogenous to both the bail amount and detention status, a large change in effect size as a result of these controls would raise concerns that the effects are being driven by variations in the magistrate's willingness to grant public defense. Alternatively, it could suggest that the effects seen are not as a result of pretrial detention per se, but rather the wealth affects of bail and the ability to hire a quality lawyer. However the inclusion of controls for having a public defender, a court-appointed attorney, or a private attorney have only trivial results on the estimates.

In Figure 5 I look at how effect sizes differ if pretrial detention is defined as spending at least one night in jail after the bail hearing, more than two weeks, more than thirty days, or still being in jail at the time of disposition. The figures show regression estimates from the IV specification where magistrate's preferences are allowed to vary over time and according to defendant characteristics. Figures 5a and b show results for the full sample, Figure 5c shows the weak-evidence sample and Figure 5d shows cases for first or second time arrestees. The outcome in Figure 5b is conviction and the outcome for all other figures is guilty pleas.

In each subplot, the effect size increases as the number of days detained increases. The standard errors increase as well. This is because the initial bail amount set by the magistrate becomes less relevant to detention status as time goes on (future judges may revise bail downward).[27] However despite the noisy estimates, the lower bound on the 95% confidence interval is far from zero in some of these specifications. Among weak-evidence crimes, the lower bound of the effect of being detained throughout the entire pretrial period on the likelihood of pleading guilty is almost 10 percentage points.

10 Conclusion

Using a natural experiment in Philadelphia where the likelihood of being detained pretrial is exogenously affected by the magistrate who presides over the bail hearing, I find that pretrial detention leads to an increase in the likelihood of being convicted, mostly by increasing the likelihood that defendants, who otherwise would have been acquitted or had their charges dropped, plead guilty. The effects are larger among first or second time arrestees

[27] I chose 'detained more than three days' as my preferred measure since it seemed a reasonable balance between strength of instrument and size of the effect.

and among crime types where evidence tends to be weak.

In Philadelphia, almost 80 percent of arrestees have bail set at $10,000 or less. These arrestees would only need to pay up to a $1000 deposit to secure their release, an amount that is likely to be had in savings or available to borrow by most middle and upper middle class Americans. Yet 60% of arrestees with a $10,000 bail are unable to pay this amount within three days and 34% remain in jail at the time of disposition.

Some argue that money bail is unconstitutional since it is so difficult to ensure that the price of bail is set proportional to one's means in a way that precludes detention based on wealth. In fact, many jurisdictions do not take ability to pay into account at all when setting bail, a practice that is clearly a violation of the Equal Protection Clause.[28] A nonprofit legal organization called Equal Justice Under Law has been filing a series of lawsuits to ensure that the defendant's ability to pay is taken into account in the setting of bail.

Yet this is not the only important legal/policy question which hinges, at least partly, on the full costs of pretrial detention to the defendant. Another as-of-yet-unanswered question is whether defendants have a right to counsel at the bail hearing. The Supreme Court has ruled that the state must pay for indigent defendants to have an attorney at all 'critical stages' of the criminal proceedings, but what exactly constitutes a critical stage is not completely clear. Many jurisdictions, including Philadelphia, do not provide counsel to indigent clients at the bail hearing.

Another outstanding question is whether detention places undue coercive pressure on defendants to plead guilty. Guilty pleas are required to be voluntary by the Due Process Clauses of the Fifth and Fourteenth Amendments and if the 'punishment' of waiting in jail until trial is worse than the penalties involved with pleading guilty then a plea may might not be considered voluntary. Finally, the Eighth Amendment prohibits 'excessive bail'. Since the full costs of bail include its effect on case outcomes, bail amounts that might have been considered reasonable when only weighing the costs of a pretrial loss of liberty against the benefits of averted crime may seem excessive when including all the downstream effects of pretrial detention.[29]

The findings presented in this paper, taken in context of these four highly policy-relevant questions, suggest several things. First they suggest that jurisdictions should move away from a means-based method of determining who is detained pretrial, as the socio-economic disparities of such a system ripple far beyond the pretrial period. Second, they suggest that

[28]The Department of Justice issued a Statement of Interest in Varden v. City of Clanton (February 2015) declaring that fixed bail bond schedules that do not take indigence into account are a violation of the Fourteenth Amendment.

[29]For a more detailed discussion of the constitutional questions which hinge at least partly on how pretrial detention affects case outcomes see (Heaton et al., 2016).

the bail hearing is indeed a critical stage, and indigent defendants should have the right to counsel. Third, they suggest that low risk defendants – for whom the 'punishment' of pretrial detention is worse than the expected punishment for the crime – should not be detained as detention may result in coerced guilty pleas. Finally, while determinations of 'excessiveness' are beyond the scope of this paper, the evidence suggests that the costs of detention are high and, if justified, the benefits must be high as well.

References

Abrams, David S. and Chris Rohlfs, "Optimal Bail and the Value of Freedom: Evidence from the Philadelphia Bail Experiment," *Economic Inquiry*, 2011, *49* (3), 750–770.

Aizer, Anna and Joseph J. Jr. Doyle, "Juvenile Incarceration, Human Caoital and Future Crime: Evidence from Randomly-Assigned Judges," *National Bureau of Economic Research Working Paper*, 2009.

Allen, Ronald J. and Larry Laudan, "Deadly Dilemmas II: Bail and Crime," *Northwestern Public Law Research Paper*, December 2008, (08-44).

Anderson, James M. and Paul Heaton, "How Much Difference Does the Lawyer Make? The Effect of Defense Counsel on Murder Case Outcomes," *Yale Law Journal*, 2012, *122* (1).

Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger, "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 1999, *14* (1), 57–67.

Ares, Charles E., Anne Rankin, and Herbert Sturtz, "The Manhattan Bail Experiment: An Interim Report on the Use of Pretrial Parole," *New York University Law Review*, 1963, *38.*

Bibas, Stephanos, "Plea Bargaining Outside the Shadow of the Trial," *Harvard Law Review*, June 2004, *117* (8).

Clark, John, Daniel Peterca, and Stuart Cameron, "Assessment of Pretrial Services in Philadelphia," Technical Report, Pretrial Justice Institute February 2011.

Cohen, Thomas H. and Brian A. Reaves, "Pre-Trial Release of Felony Defendants in State Court," Technical Report, Bureau of Justice Statistics Special Report November 2007.

Devers, Lindsey, "Plea and Charge Bargaining: Research Summary," BJA Report January 2011.

Free, Marvin D. Jr., "Bail and Pretrial Release Decisions," *Journal of Ethnicity in Criminal Justice*, 2004, *2* (4), 23–44.

Goldkamp, John S., "The Effects of Detention on Judicial Decisions: A Closer Look," *The Justice System Journal*, 1980, *5* (3), 234–257.

Heaton, Paul, Sandra Mayson, and Megan Stevenson, "The Downstream Criminal Justice Consequences of Pretrial Detention," Working Paper 2016.

Kahneman, Amos Tversky Daniel, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–291.

Kling, Jeffrey R., "Incarceration Length, Employment, and Earnings," *American Economic Review*, June 2006, *96* (3), 863–876.

Loeffler, Charles E., "Does Imprisonment Alter the Life Course? Evidence on Crime and Employment From a Natural Experiment," *Criminology*, 2013, *51* (1), 137–166.

Lowenkamp, Christopher T., Marie VanNostrand, and Alexander Holsinger, "Investigating the Impact of Pretrial Detention on Sentencing Outcomes," Technical Report, Laura and John Arnold Foundation 2013.

Mayson, Sandra and Megan Stevenson, "Cost of Crime Versus the Cost of Doing Time: Is Pretrial Detention 'Cost-Justified' When the Loss To Legally Innocent Detainees is Included?," Working Paper 2016.

Minton, Todd D. and Zhen Zeng, "Jail Inmates at Midyear 2014," Technical Report, Bureau of Justice Statistics Bulletin June 2015.

Mueller-Smith, Michael, "The Criminal and Labor Market Impacts of Incarceration: Identifying Mechanisms and Estimating Household Spillovers," Working Paper August 2015.

Oleson, J.C., Christopher T. Lowenkamp, Timothy P. Cadigan, Marie VanNostrand, and John Wooldredge, "The Effect of Pretrial Detention on Sentencing in Two Federal Districts," *Justice Quarterly*, 2014.

ONDC, ""FAFSA Facts"," Technical Report, Office of National Drug Control Policy U.S. Department of Education.

Phillips, Mary T., "Pretrial Detention and Case Outcomes, Part 1: Nonfelony Cases," Final Report, New York Criminal Justice Agency, Inc. November 2007.

_ , "Bail, Detention and Felony Case Outcomes," Research Brief, New York Criminal Justice Agency, Inc. September 2008.

Rankin, Anne, "The Effect of Pretrial Detention," *New York University Law Review*, 1964, *39.*

Sacks, Meghan and Alissa R. Ackerman, "Pretrial detention and guilty pleas: if they cannot afford bail they must be guilty," *Criminal Justice Studies*, 2012, *25* (3), 265–278.

Schlesinger, Traci, "Racial and ethnic disparity in pretrial criminal processing," *Justice Quarterly*, 2005, *22* (2), 170–192.

SHRM, ""Society for Human Resource Management Survey Findings: Background Checking - The Use of Criminal Background Checks in Hiring Decisions"," Technical Report, The Sentencing Project July 2012.

Shubik-Richards, Claire and Don Stemen, "Philadelphia's Crowded, Costly Jails: The Search for Safe Solutions," Technical Report, Pew Charitable Trusts Philadelphia Research Inititiative May 2010.

Tartaro, Christine and Christopher M. Sedelmaier, "A tale of two counties: the impact of pretrial release, race, and ethnicity upon sentencing decisions," *Criminal Justice Studies*, 2009, *22* (2), 203–221.

Tella, Rafael Di and Ernesto Schargrodsky, "Criminal Recidivism after Prison and Electronic Monitoring," *Journal of Political Economy*, 2013, *121* (1), 28 – 73.

Thacher, David, "The Rise of Criminal Background Screening in Rental Housing," *Law & Social Inquiry*, 2008, *33* (1), 5–30.

Uggen, Christopher, Sarah Shannon, and Jeff Manza, ""State-Level Estimates of Felon Disenfranchisement in the United States, 2010"," Technical Report, The Sentencing Project July 2012.

Williams, Marian R., "The Effect of Pretrial Detention on Imprisonment Decisions," *Criminal Justice Review*, 2003, *28* (2), 299–316.

Wooldredge, John, "Distinguishing Race Effects on Pre-Trial Release and Sentencing Decisions," *Justice Quarterly*, 2012, *29* (1), 41–75.

Table 1: Summary statistics

	Released	Detained	Total
Age	32.9	32.0	32.5
Male	0.79	0.89	0.83
Caucasian	0.30	0.26	0.29
African-American	0.53	0.65	0.57
Missing race	0.15	0.05	0.11
Number of prior cases	4.14	6.25	4.97
Has felony charge	0.36	0.75	0.51
Number of charges	4.95	6.71	5.64
Bail	\$3,345	\$63,336	\$26,877
Bail=\$0	0.54	0.01	0.33
Detained>3 days	0	1	0.41
Never released	0.00	0.58	0.23
All charges dropped	0.48	0.48	0.48
Case went to trial	0.32	0.19	0.27
Not guilty on all charges	0.03	0.03	0.03
Guilty of at least one charge	0.49	0.49	0.49
Pled guilty to at least one charge	0.21	0.33	0.26
Court fees charged	\$386	\$211	\$317
Sentenced to incarceration	0.18	0.34	0.24
Days of incarc. sentence	94	585	292
Days before elig. for parole	42	331	159
Observations	197,775	133,840	331,615

Note: Summary statistics are presented for those who are released within three days of the bail hearing (Column 1), those who are detained for more than three days after the bail hearing (Column 2) and the entire sample (Column 3). All offense variables refer to the charges present at the time of the bail hearing. The statistic shown is the mean and, unless otherwise indicated, variables are dummies where 1 indicates the presence of a characteristic. Age is measured in years, those marked "Number..." are count variables, and those expressed in dollar amounts are currency. The bottom two rows refer to the maximum number of days of the incarceration sentence and the minimum number of days before the defendant is eligible for parole. The sentence is coded as zero if the defendant did not receive an incarceration sentence.

Table 2: Main offenses

DUI, 1st offense	0.065
Prostitution	0.020
Shoplifting	0.042
Small amount marijuana	0.022
Illegal firearms (F2 and F3)	0.036
Drug possession	0.14
Buying drugs	0.053
Selling drugs	0.13
Car theft	0.021
Theft	0.042
Robbery	0.073
Burglary	0.046
Murder	0.021
Vandalism	0.011
Simple assault	0.065
Aggravated assault	0.091
Observations	331615

Note: The statistic shown is the fraction of the overall sample whose lead charge is as listed. Crime types are listed according to strength of evidence; crime types at the top of the list tend to have stronger evidence than those at the top of the list. Strength of evidence is measured both by a poll of criminal justice lawyers and by average conviction rate. The horizontal line separates those placed in the 'weak-evidence' category and those placed in the 'strong-evidence' category

Table 3: How do race and neighborhood wealth relate to bail amount and pretrial detention?

	(1) Log bail amount	(2) Pretrial detention
African-American	0.00197	0.0278****
	(0.0125)	(0.00178)
Log income (average per zip code)	-0.0451***	-0.0217****
	(0.0144)	(0.00205)
Observations	251236	251236
R^2	0.584	0.336
Demographic controls	Y	Y
Criminal history controls	Y	Y
Offense controls	Y	Y

Standard errors in parentheses

Heteroskedastic-Robust Standard Errors

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows the how race and neigborhood wealth correlate with bail amount and pretrial detention status, after controlling for offense, criminal history, age, gender and time. Only the subset of defendants for whom zipcode and race information was available were included

Table 4: Covariate balance across magistrates

Summary statistics for defendant characteristics	F statistic	Empirical p value
Predicted likelihood of pretrial detention	1.84	0.96
Predicted likelihood of pleading guilty	2.59	0.56
Predicted likelihood of conviction	1.91	0.58

Note: The predicted likelihoods described in the left-most columns are the predicted values from a regression of pretrial detention, guilt, and conviction, respectively, on offense, criminal history, demographics and time controls. The F statistics are the F statistics in a test of joint significance of eight magistrate dummies when regressing the predicted values on the magistrate dummies and controls for the time and date of the bail hearing. The empirical p values show the likelihood of seeing an F statistic as big or bigger if the magistrate seen was due to chance; this is calculated as the fraction of 'false' F statistics as big or bigger in a permutation test.

Table 5: How does pretrial detention affect conviction rates and guilty pleas?

Panel A: Full sample (IV)	Conviction (mean dep. var.= 0.49)					
	(1)	(2)	(3)	(4)	(5)	(6)
Pretrial detention	0.166**	0.181***	0.252***	0.122***	0.0871**	0.0665**
	(0.0734)	(0.0653)	(0.0794)	(0.0411)	(0.0369)	(0.0293)
Panel B: Full sample (IV)	**Guilty pleas (mean dep. var.=0.25)**					
	(1)	(2)	(3)	(4)	(5)	(6)
Pretrial detention	0.124**	0.174***	0.175**	0.104***	0.0582*	0.0531**
	(0.0617)	(0.0561)	(0.0715)	(0.0364)	(0.0329)	(0.0265)
Magistrate X 3 time periods		Y	Y	Y	Y	Y
Magistrate X top 5 crimes				Y	Y	Y
Magistrate X crim. history					Y	Y
Magistrate X demographics					Y	Y
Magistrate X top 16 crimes						Y
Time controls	Y	Y	Y	Y	Y	Y
Covariates			Y	Y	Y	Y
Observations	331615	331615	331615	331615	331615	331615
First stage F	34.87	19.53	26.09	21.87	14.75	11.65
Number of instruments	8	8	19	59	107	203
Mean indep. var	0.41	0.41	0.41	0.41	0.41	0.41

Standard errors in parentheses

Heteroskedastic-Robust Standard Errors

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: The dependent variable in Panels A and B respectively is a dummy equal to one if the defendant is convicted on at least one charge and a dummy equal to one if the defendant pled guilty on at least one charge. The instruments in the first two columns are the eight magistrate dummies; in the subsequent columns the instruments include interactions between the magistrate dummies and three time period fixed effects, the five most common crime types, a variety of criminal history variables, defendant demographics, and the remainder of the 16 most common crime types as shown in Table 2. The first two columns control only for the time and date of the bail hearing, all subsequent columns include the full set of controls as described in Footnote 21. The F statistic on the exogenous first stage instruments is listed at the bottom, as are the number of instruments used in that specification and the mean of the independent variable. A linear jackknife instrumental variables regression is used. The Rˆ2 is not reported due to difficulties of interpreting this statistic in an IV regression.

Table 6: Full sample results - jackknife IV and OLS

	(1) Conv-iction	(2) Guilty plea	(3) Court Fees	(4) Incarc-eration	(5) Max days	(6) Min days
Panel A: Full sample (IV)						
Pretrial detention	0.0665**	0.0531**	128.7****	0.0193	119.3	140.6**
	(0.0293)	(0.0265)	(33.59)	(0.0251)	(73.40)	(61.86)
Panel B: Full sample (OLS)						
Pretrial detention	0.0355****	0.0558****	-103.0****	0.1000****	136.9****	69.88****
	(0.00197)	(0.00181)	(2.621)	(0.00166)	(3.430)	(2.518)
Observations	331615	331615	331615	331615	331613	331613
First stage F	11.65	11.65	11.65	11.65	11.65	11.65
Mean dep. var.	0.49	0.25	$312	0.24	292	155
Mean indep. var.	0.41	0.41	0.41	0.41	0.41	0.41

Standard errors in parentheses

Heteroskedastic-Robust Standard Errors

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows how pretrial detention affects various case outcomes using both a jackknife IV regression (Panel A) and an OLS regression (Panel B). The outcome variables are dummies for being convicted/pleading guilty, total non-bail court fees in dollars, a dummy for whether or not the defendant receives an incarceration sentence, the maximum days of that incarceration sentence and the minimum days the defendant must serve before being eligible for parole. In all of the IV specifications magistrate preferences are allowed to vary across three time periods and according to offense, criminal history and demographics of defendants. The F statistic on the first stage of the jackknife IV are shown in the sub-panel, as are the means of the dependent and independent variables. All regressions include the full set of controls as described in Section 6.

Table 7: Comparing results for misdemeanors and felonies

	(1) Conv-iction	(2) Guilty plea	(3) Court Fees	(4) Incarc-eration	(5) Max days	(6) Min days
Panel A: Misdemeanors (IV)						
Pretrial detention	0.0850**	0.0684**	93.48**	0.0851***	66.00***	30.68**
	(0.0366)	(0.0300)	(37.85)	(0.0279)	(21.64)	(12.13)
Panel B: Misdemeanors (OLS)						
Pretrial detention	0.0190****	0.0515****	-13.34****	0.0513****	39.23****	19.62****
	(0.00298)	(0.00249)	(3.081)	(0.00213)	(2.087)	(1.401)
Observations	163125	163125	163125	163125	163124	163124
First stage F	12.82	12.82	12.82	12.82	12.82	12.82
Mean dep. var.	0.50	0.16	$351	0.16	48	19
Mean indep. var.	0.23	0.23	0.23	0.23	0.23	0.23
Panel C: Felonies (IV)						
Pretrial detention	0.0598	0.0545	136.7**	-0.0214	147.5	179.0
	(0.0433)	(0.0414)	(53.66)	(0.0398)	(135.3)	(115.2)
Panel D: Felonies (OLS)						
Pretrial detention	0.0514****	0.0589****	-174.2****	0.134****	194.7****	98.10****
	(0.00266)	(0.00259)	(4.023)	(0.00244)	(5.583)	(4.053)
Observations	168490	168490	168490	168490	168489	168489
First stage F	7.92	7.92	7.92	7.92	7.92	7.92
Mean dep. var.	0.47	0.35	$274	0.32	528	294
Mean indep. var.	0.58	0.58	0.58	0.58	0.58	0.58

Standard errors in parentheses

Heteroskedastic-Robust Standard Errors

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows effect sizes in misdemeanor crimes (Panels A and B) and felonies (Panel C and D). The outcome variables are dummies for being convicted/pleading guilty, total non-bail court fees in dollars, a dummy for whether or not the defendant receives an incarceration sentence, the maximum days of that incarceration sentence and the minimum days the defendant must serve before being eligible for parole. In all IV specifications magistrate preferences are allowed to vary across three time periods and according to offense, criminal history and demographics of defendants. The F statistic on the first stage of the jackknife IV are shown in the sub-panel, as are the means of the dependent and independent variables. All regressions include the full set of controls as described in Footnote 21.

Table 8: Effect sizes by strength of evidence

	(1) Conv-iction	(2) Guilty plea	(3) Court Fees	(4) Incarc-eration	(5) Max days	(6) Min days
Panel A: Weak-evidence crimes (IV)						
Pretrial detention	0.0415	0.0735*	-42.86	-0.00348	516.3****	541.1****
	(0.0421)	(0.0383)	(44.04)	(0.0354)	(141.7)	(124.1)
Panel B: Weak-evidence crimes (OLS)						
Pretrial detention	0.0983****	0.0894****	-67.93****	0.131****	171.3****	83.71****
	(0.00316)	(0.00285)	(3.797)	(0.00250)	(7.089)	(5.703)
Observations	122742	122742	122742	122742	122741	122741
First stage F	9.10	9.10	9.10	9.10	9.10	9.10
Mean dep. var.	0.35	0.25	$158	0.20	466	276
Mean indep. var.	0.56	0.56	0.56	0.56	0.56	0.56
Panel C: Strong-evidence crimes (IV)						
Pretrial detention	0.0308	0.0277	206.1****	0.0132	3.994	0.857
	(0.0415)	(0.0369)	(49.26)	(0.0357)	(37.54)	(17.61)
Panel D: Strong-evidence crimes (OLS)						
Pretrial detention	0.00707**	0.0379****	-111.3****	0.0832****	98.36****	47.45****
	(0.00289)	(0.00273)	(4.167)	(0.00262)	(3.568)	(1.809)
Observations	165488	165488	165488	165488	165488	165488
First stage F	13.29	13.29	13.29	13.29	13.29	13.29
Mean dep. var.	0.60	0.27	$435	0.27	187	88
Mean indep. var.	0.27	0.27	0.27	0.27	0.27	0.27

Standard errors in parentheses

Heteroskedastic-Robust Standard Errors

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: Panels A and B show effect sizes in case types where evidence tends to be relatively weak (murder, aggravated assault, simple assault, vandalism, burglary, theft, robbery, car theft) and Panels C and D show case types where evidence tends to be relatively strong (DUI, drug possession, illegal firearms, drug sale, drug purchase, prostitution, shoplifting). The outcome variables are dummies for being convicted/pleading guilty, total non-bail court fees in dollars, a dummy for whether or not the defendant receives an incarceration sentence, the maximum days of that incarceration sentence and the minimum days the defendant must serve before being eligible for parole. In all specifications magistrate preferences are allowed to vary across three time periods and according to offense, criminal history and demographics of defendants. The F statistic on the first stage of the jackknife IV are shown in the sub-panel, as are the means of the dependent and independent variables. All regressions include the full set of controls as described in Footnote 21.

Table 9: Comparing results across defendant characteristics

	(1) Conv-iction	(2) Guilty plea	(3) Court Fees	(4) Incarc-eration	(5) Max days	(6) Min days
Panel A: White defendants (IV)						
Pretrial detention	0.0700	0.0265	52.43	-0.0253	146.7	186.0*
	(0.0596)	(0.0557)	(75.81)	(0.0541)	(129.6)	(102.1)
Observations	93937	93937	93937	93937	93937	93937
Mean dep. var.	0.55	0.29	$361	0.27	254	124
Panel B: Black defendants (IV)						
Pretrial detention	0.0698*	0.0355	136.4***	-0.00125	75.48	116.2
	(0.0400)	(0.0359)	(45.19)	(0.0342)	(111.3)	(94.53)
Observations	191200	191200	191200	191200	191199	191199
Mean dep. var.	0.49	0.25	$296	0.25	357	196
Panel C: Defendants under 30 (IV)						
Pretrial detention	0.0487	0.0909	87.38	-0.0106	296.2	286.8
	(0.0658)	(0.0598)	(79.66)	(0.0573)	(214.3)	(187.3)
Observations	167392	167392	167392	167392	167391	167391
Mean dep. var.	0.47	0.27	$304	0.24	348	193
Panel D: Defendants over 30 (IV)						
Pretrial detention	0.0748**	0.0553*	174.0****	0.0254	18.56	55.91
	(0.0359)	(0.0326)	(40.90)	(0.0308)	(74.48)	(60.45)
Observations	164194	164194	164194	164194	164193	164193
Mean dep. var.	0.51	0.25	$320	0.24	235	117
Panel E: First or second time arrestees (IV)						
Pretrial detention	0.175**	0.122*	-40.23	-0.0333	194.0	300.5
	(0.0751)	(0.0686)	(95.65)	(0.0588)	(225.8)	(192.3)
Observations	113932	113932	113932	113932	113930	113930
Mean dep. var.	0.41	0.22	$313	0.17	202	108
Panel F: Defendants with two or more prior arrests (IV)						
Pretrial detention	0.0541*	0.0387	170.6****	0.0736***	185.8**	180.6***
	(0.0317)	(0.0286)	(36.03)	(0.0282)	(75.66)	(63.08)
Observations	217683	217683	217683	217683	217683	217683
Mean dep. var.	0.53	0.28	$312	0.27	339	180

Standard errors in parentheses

Heteroskedastic-Robust Standard Errors

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows effect sizes among black defendants (Panel A), white defendants (Panel B), defendants under 30 (Panel C), defendants over 30 (Panel D), first or second time arrestees (Panel E) and defendants with two or more prior arrests (Panel F). In all specifications magistrate preferences are allowed to vary across three time periods and according to offense, criminal history and demographics of defendants.

Table 10: Robustness checks

Panel A: Full sample, controlling for magistrate fixed effects (IV)						
	(1) Conviction	(2) Guilty plea	(3) Court Fees	(4) Incarceration	(5) Max days	(6) Min days
Pretrial detention	0.0371	0.0413	106.0***	0.00242	85.32	160.0**
	(0.0326)	(0.0298)	(37.28)	(0.0283)	(92.73)	(79.78)
Observations	331615	331615	331615	331615	331613	331613

Panel B: Full sample, controlling for attorney type (IV)						
	(1) Conviction	(2) Guilty plea	(3) Court Fees	(4) Incarceration	(5) Max days	(6) Min days
Pretrial detention	0.0731**	0.0587**	124.3****	0.0258	114.4	134.9**
	(0.0287)	(0.0260)	(33.56)	(0.0247)	(73.05)	(61.70)
Observations	331615	331615	331615	331615	331613	331613

Standard errors in parentheses

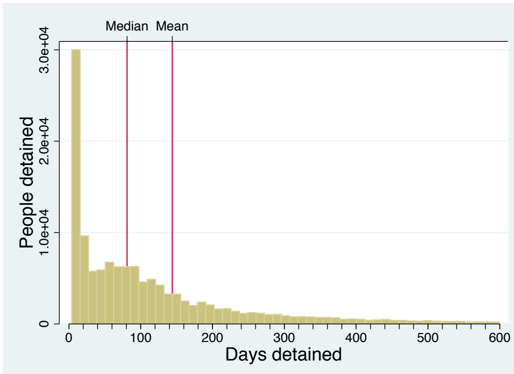
Heteroskedastic-Robust Standard Errors

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

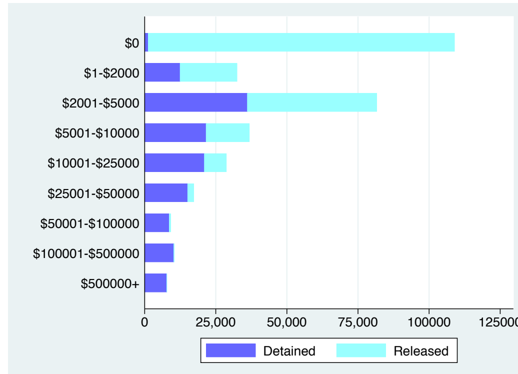
Note: This table presents robustness checks for the main results. Panel A includes magistrate fixed effects as controls; the effects of pretrial detention are thus identified solely off of within-magistrate variation in detention rates. Panel B includes controls for attorney type: public defender, court-appointed attorney, and private. The outcome variables are dummies for being convicted/pleading guilty, total non-bail court fees in dollars, a dummy for whether or not the defendant receives an incarceration sentence, the maximum days of that incarceration sentence and the minimum days the defendant must serve before being eligible for parole. In all specifications magistrate preferences are allowed to vary across three time periods and according to offense, criminal history and demographics of defendants. All regressions include the full set of controls as described in Footnote 21.

Figure 1: Descriptive graphs

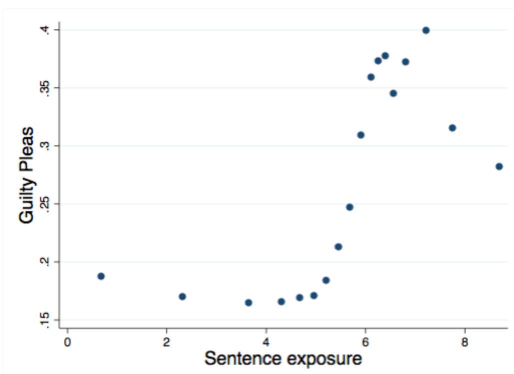
(a) Days detained pretrial, conditional on being detained more than three days



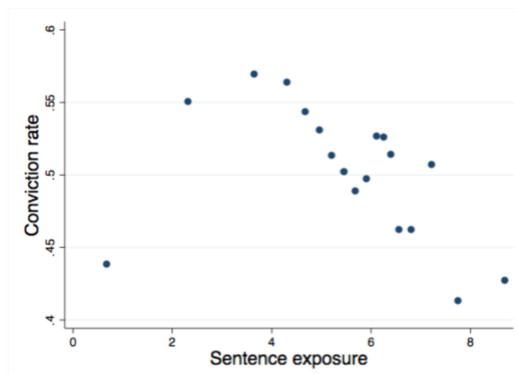
(b) Bail amounts and detention status



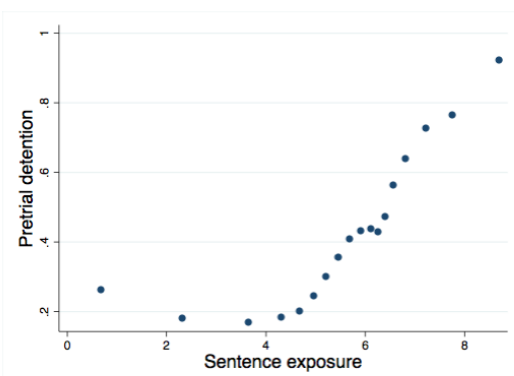
(c) Likelihood of pleading guilty at different levels of sentence exposure



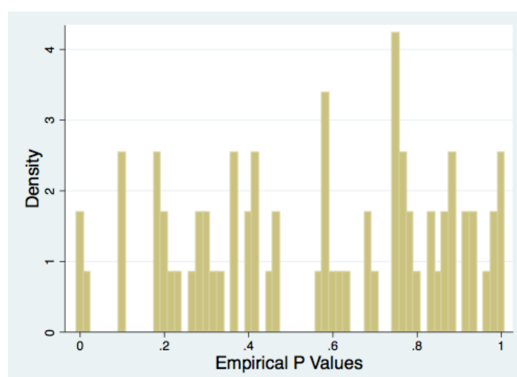
(d) Likelihood of conviction at different levels of sentence exposure



(e) Likelihood of being detained pretrial at different levels of sentence exposure



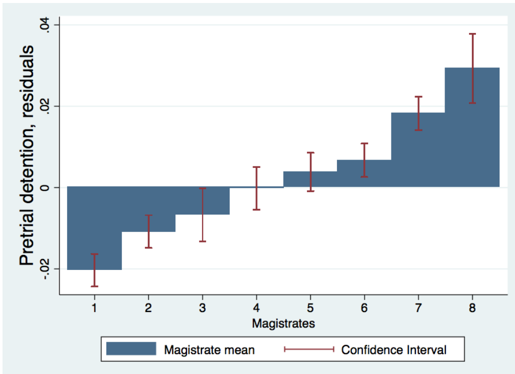
(f) Distribution of 69 'empirical p values' from permutation test



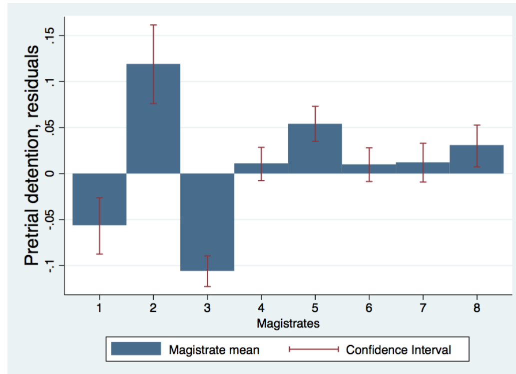
Note: Figure 1a shows the number of people detained/released at various levels of bail. Figure 1b shows the number of days detained pretrial for those who are detained for more than three days. Figures 1c, d and e are a binned scatterplots showing the fraction pleading guilty, convicted, and detained pretrial at various levels of sentence exposure. 'Sentence exposure' is a log transform of the predicted value from a regression of days of the incarceration sentence on offense, criminal history, demographics and time controls, with the sample limited to those who were found guilty at trial. Figure 1f is a histogram of the 'empirical p values' of 69 different permutation tests to evaluate covariate balance across magistrates.

Figure 2: Average detention rates by magistrate for different offense types

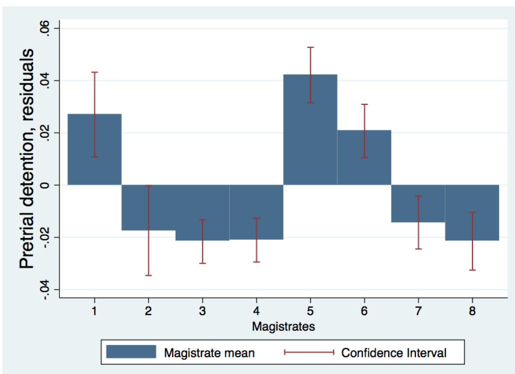
(a) All cases (n=331,615, F=38.56)



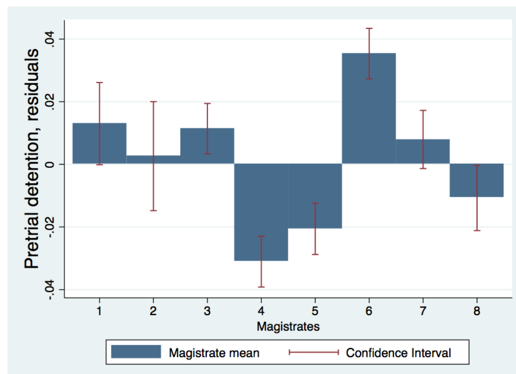
(b) Shoplifting (n=15,775, F=31.82)



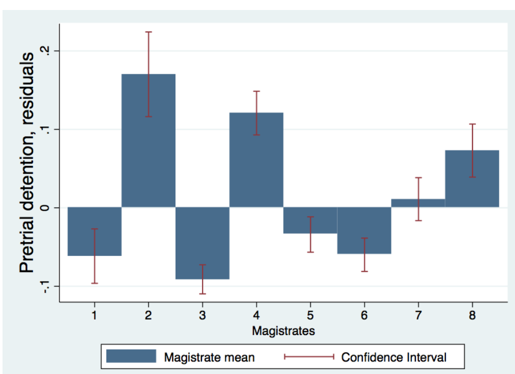
(c) DUI, 1st offense (n=25,850, F=25.88)



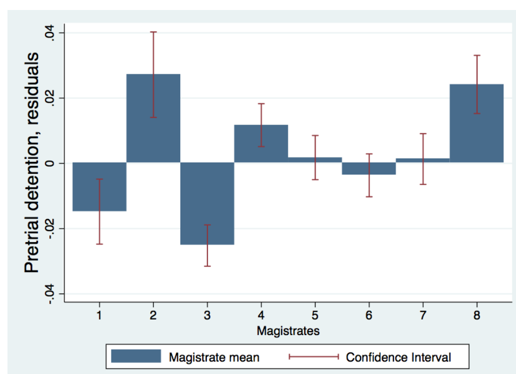
(d) Simple assault (n=85,396, F=24.93)



(e) Prostitution (n=6,529, 42.14)



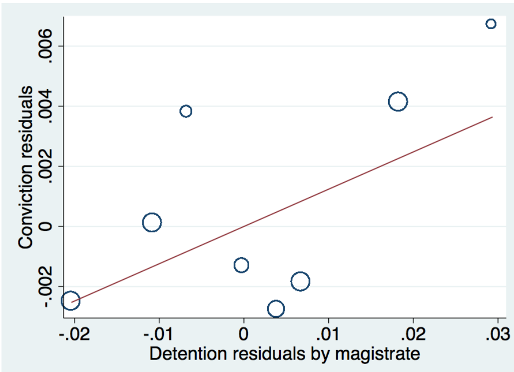
(f) Drug possession (n=109,042, F=18.61)



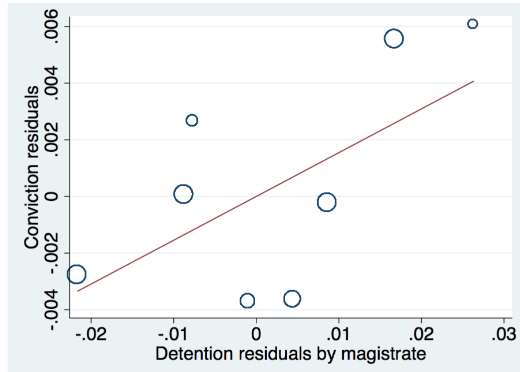
Note: This figure shows pretrial detention rates by magistrate over the whole sample (Figure 2a), and for different offense categories (Figure 2b-f). The numbers 1 through 8 delineate the different magistrates. The y axes show the residuals from a regression of pretrial detention on time controls. The error bars indicate 95% confidence intervals for the mean. n indicates the number of observations per category, and the F statistic refers to a joint F statistic on the eight magistrate dummies when regressing pretrial detention on the magistrate dummies and time controls. The numbering of the magistrates is consistent across all samples.

Figure 3: Visual IV

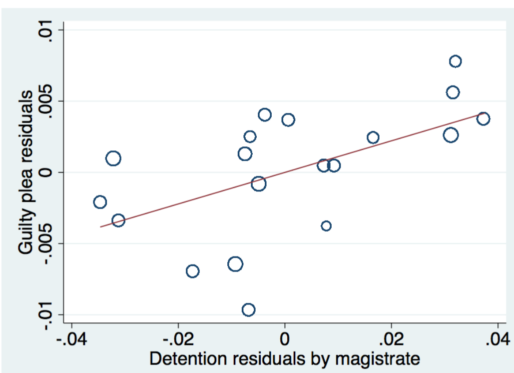
(a) Full sample – conviction rates and pretrial detention are residualized over time controls



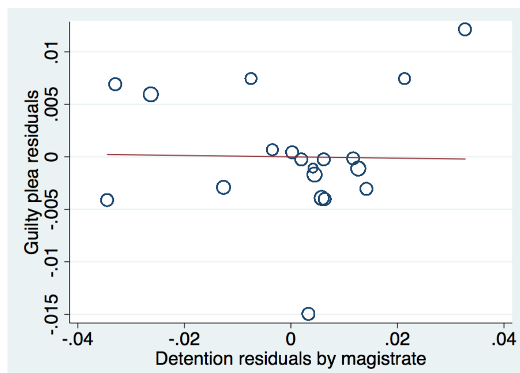
(b) Full sample – conviction rates and pretrial detention are residualized over time controls, offense, criminal history and demographics



(c) Weak-evidence crimes – guilty plea rate and pretrial detention are residualized over time controls, offense, criminal history and demographics

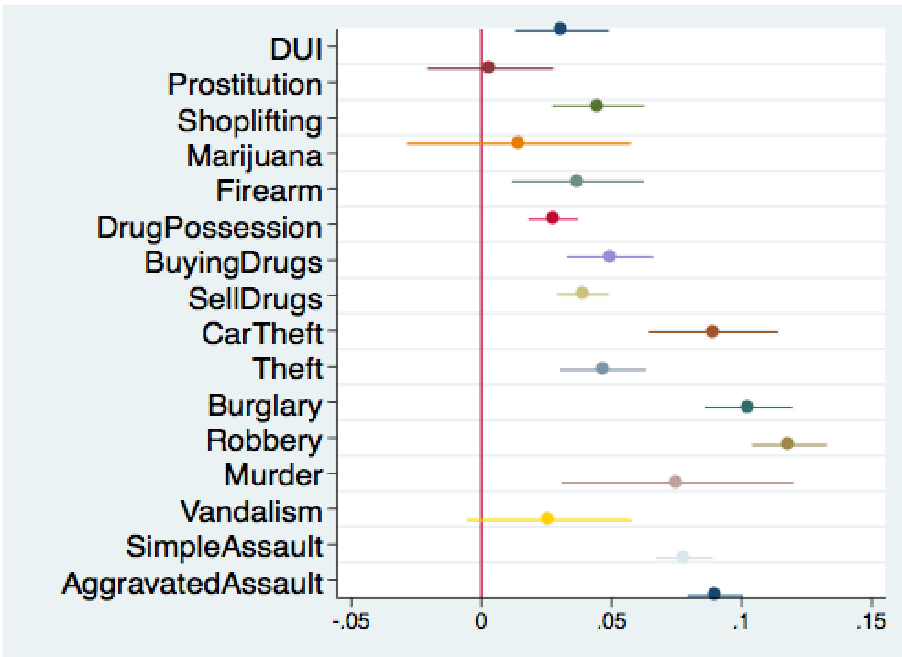


(d) Strong-evidence crimes – guilty plea rate and pretrial detention are residualized over time controls, offense, criminal history and demographics



Note: The y and x axes in Figure 3a show the residuals from a regression of a dummy for conviction and pretrial detention (respectively) on controls for the time and date of the bail hearing. Figure 3b is the same, except conviction and detention have been residualized over offense, criminal history and demographic covariates as well as time controls. The circles in Figures 3a-b show the average detention and conviction residuals for each magistrate; the size of the circle is proportional to the number of cases seen by that magistrate. The y and x axes in Figures 3c-d are residuals from a regression of pleading guilty on offense, criminal history, demographic and time controls. Figure 3c shows the weak-evidence sample and Figure 3d shows the strong-evidence sample. Here the circles represent the average detention and guilty plea residuals per-magistrate-per-time period. There are three time periods as separated by February 23, 2009 and February 23, 2011.

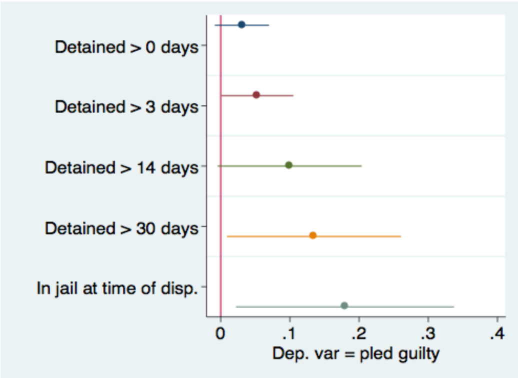
Figure 4: OLS estimates of the impact pretrial detention has on guilty pleas within different offense categories, ordered by strength of evidence



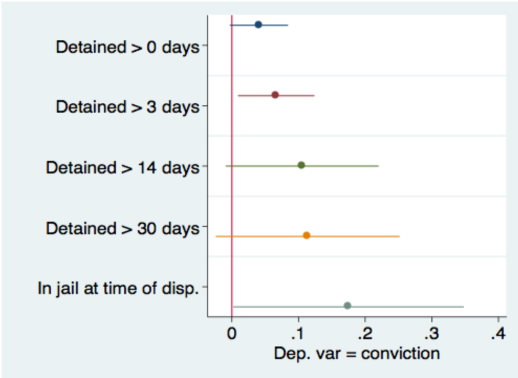
Note: The above coefficient plots show the OLS estimates of the impact pretrial detention has on guilty pleas for different offense, as labeled on the left. The offenses are ordered according to the strength of evidence that tends to be present for different case types; the offenses on the top tend to have the strongest evidence. Each dot represents the estimated coefficient on pretrial detention, the line represents the 95% confidence interval.

Figure 5: Coefficient plots showing the impact of pretrial detention on case outcomes using various definitions of 'pretrial detention'

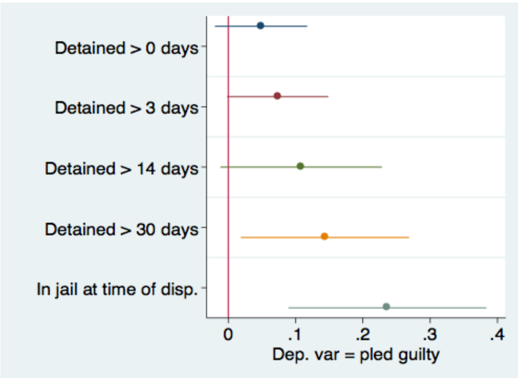
(a) Full sample



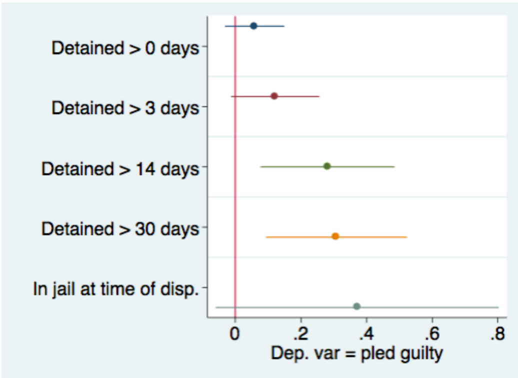
(b) Full sample



(c) Weak-evidence crimes



(d) First or second arrest



Note: The above coefficient plots show jackknife IV regression results where the endogenous independent variable is defined as being detained at least one day, greater than three days, greater than 14 days, greater than 30 days, or until the time of disposition. The top two plots show results from the full sample, the bottom left plot shows results from the 'weak-evidence' sample and the bottom right shows results from the sample for first or second time arrestees. The instruments in all specifications are the magistrate dummies interacted with offense, criminal history, race, gender, and three time periods; the full set of controls are included. The dot shows the magnitude of the coefficient estimate as indicated on the x axis and the line indicates a 95% confidence interval.

Article 2: Introductory Note

Using rich data on youths in juvenile correctional facilities, I conduct a series of tests of peer influence on future crime motivated by three mechanisms: criminal skill transfer, the formation of new criminal networks, and the social contagion of crime-oriented attitudes and behavioral habits. Identifying peer influence off of natural variation in small cohorts within the same facility, I find evidence consistent with the social contagion mechanism: exposure to peers who come from unstable homes and who have behavioral/emotional problems leads to a large increase in crime after release, as well as an increase in crime-oriented non-cognitive factors. I also find evidence consistent with persistent network formation, but only in settings which unite youths from the same local area. Multiple tests of the identifying assumptions support the causal argument.

Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails

1 Introduction

A variety of papers show evidence that peers are influential on crime, but almost all are reduced form studies where the mechanisms of influence remain inside the black box (Glaeser et al., 1995; Jacob and Lefgren, 2003; Kling et al., 2007; Chen and Shapiro, 2007; Bayer et al., 2009; Ouss, 2011; Drago and Galbiati, 2012; Lerman, 2013; Billings et al., 2013; Damm and Gorinas, 2013). Without knowledge of the mechanisms through which peers influence crime it is difficult to design effective policy or to extrapolate the results of a study beyond the time and place in which it was conducted. Using a dataset on incarcerated youths that is unprecedented in terms of detail and size, I conduct a series of tests of peer influence that are designed to not only identify causal peer influence, but also distinguish between three potential mechanisms: criminal skill transfer, persistent network formation, and the social contagion of crime-oriented attitudes and behavioral habits. I find that the peer cohort in incarceration facilities has a strong effect on post-release crime. The evidence is consistent with the social contagion mechanism, and would be difficult to reconcile with either of the other two mechanisms. Separately I find evidence supportive of the persistent network formation mechanism in correctional settings that unite youths from the same local area. Although I cannot rule it out, I do not find evidence that supports the skill transfer mechanism.

While this research has ramifications beyond its specific setting, juvenile incarceration is an important area of research in its own right. Although juvenile violent crime rates are only marginally higher than in other countries, the United States incarcerates five times more adolescents than any other country (Hazel, 2008). At $88,000 per year the cost of incarcerating one juvenile is roughly five times the cost of room, board and tuition at a public university (Mendel, 2011). Approximately 75% of teenagers released from a juvenile incarceration facility are rearrested within two years; approximately 30% are re-incarcerated within that same time window (Mendel, 2011). An instrumental variables approach that leverages random assignment to judges shows that juvenile incarceration leads to an increase in both adult imprisonment and a decrease in high school graduation rates (Aizer and Doyle, 2013). One proposed explanation has to do with social dynamics of incarceration: isolation from the rest of society and around-the-clock exposure to criminally experienced and life-hardened peers.

The most well known paper in this field is by Bayer et al. (2009) which shows that exposure to peers with experience in a particular crime category leads to an increase in the propensity to commit that particular crime after release from incarceration, but only for those who already have experience in that crime category. While this paper broke ground in showing evidence of peer influence in incarceration facilities, it raises many unanswered questions. The effect was found in only about half of the skilled crimes they tested, and also for crimes of passion (assault, sex offense). That the effect would be limited to only those who already have experience in that particular crime is surprising. The authors take no stand as to the mechanism behind these results, but without understanding the mechanism the theoretical and policy implications of this paper are unclear.[1]

My research suggests that the most important mechanism of peer influence in juvenile incarceration facilities is a psychological one. I find that the peers who are most influential are not the ones with the most criminal skill but rather those who come from unstable homes and who have emotional and behavioral problems. Exposure to these peers leads to an increase in crime, but also an increase in crime-oriented non-cognitive factors such as aggression or anti-societal attitudes. This effect persists despite controlling for the criminal experience of the peer cohort as well as a variety of other peer characteristics. Furthermore, this effect is found in institutions where geographical constraints prevent peers from having much physical interaction after release.

Many papers have established the importance of non-cognitive factors in predicting education, labor force, or criminal outcomes, but how these factors are developed remains largely unknown (Bowles et al., 2001; Heckman and Rubinstein, 2001; Nagin and Pogarsky, 2004; Blanden et al., 2007; Duckworth and Peterson, 2007). One contribution of the paper is to show the role that peers play in the formation of non-cognitive characteristics. I show that peer influence over these characteristics lasts at least eight months after contact has ceased, and that these newly acquired traits are strongly predictive of criminal activity. While folk wisdom has long held that peers are influential over non-cognitive traits, showing this to be

[1] Several papers have attempted to replicate these findings with mixed results. Damm and Gorinas (2013) do not generally find an effect, with the exception being drug crimes if the peer group is defined to be those of similar ages or ethnicities. Ouss (2011) finds an effect both for skilled crimes and violence, with previous experience unimportant. However the variation in peer exposure in this setting is not exogenous, so it is unclear whether this is a causal effect.

true is difficult: it requires a credibly exogenous source of peer variation in a setting where the social exposure is intensive enough to shift deep-rooted behaviors and attitudes – and where such characteristics are actually measured. Prior research is extremely limited.

I identify peer influence using natural variation in small cohorts within the same correctional facility. The identifying assumption is that conditional on facility and time-of-release fixed effects, the variation in peer characteristics is as-good-as-random. I test this assumption in a variety of ways, and the results consistently support the causal argument. Following an 'exogenous effects' model, peer influence is identified using variation in pre-determined peer characteristics (Manski, 1993).

Each mechanism of peer influence generates testable empirical predictions about the *type* of peers who will be influential, the *settings* in which the mechanism is most likely to operate, and the *post-release behaviors* the mechanism will generate. My strategy for disentangling the mechanisms relies on a multi-pronged approach: I consider a mechanism to be supported by the evidence if there are multiple data points that are consistent with the predictions of that mechanism and inconsistent with competing explanations. I use data compiled from the Florida's Department of Juvenile Justice (FDJJ), Department of Law Enforcement, and Department of Corrections that spans all youths released from a juvenile correctional program in that state between July 2006 and July 2011. Besides crime, correctional facility, and demographic information, the data set includes 136 variables describing the youths' home, family, school and social lives. It also includes non-cognitive measures taken both before and after the incarceration period.

Although I find some evidence of the network formation mechanism in the incarceration setting, it is limited by geographical constraints. I perform a supplementary analysis on day treatment facilities (alternative schools for delinquent youth), a setting that is more conducive to the network formation mechanism since it unites youths from the same local area. Consistent with the network formation mechanism, I find that exposure to peers with lots of criminal experience or criminal connections leads to an increase in post-release crime and gang affiliation, but only if these peers live close enough to one another to enable physical interaction after release. I do not find evidence of the social contagion mechanism in this less-intensive environment. Currently, jurisdictions all over the country are moving away from incarceration and towards alternative correctional practices like the day treatment centers. By providing evidence of how peer effects work in different settings this paper adds to the discussion about the tradeoffs between types of correctional programs.

Section 2 discusses the data, assignment to facilities, and peer measures. Section 3 tests the identifying assumptions. Section 4 describes the three proposed mechanisms of peer influence, their testable predictions, and the results of these tests. Section 5 explores heterogeneity and functional form in the social contagion effect and discusses policy implications. I estimate that re-sorting the incarcerated youth so that the lower risk boys are separated from the higher risk ones would result in a 18% net decrease in the one year re-incarceration rate.[2] With yearly savings of $5.6 million in incarceration costs alone, increased funding could be provided to target the specific needs of the higher risk boys who would be adversely affected by such a shift. Cognitive behavioral therapy and other evidence-based practices that target non-cognitive skills may be effective in mitigating negative peer influence or even spawning positive peer externalities (Landenberger and Lipsey, 2005). Section 6 concludes.

2 Data, facility assignment, and peer measures

2.1 Data and descriptive statistics

The primary source of data is the Juvenile Justice Information System, which is the central administrative databank for the Florida Department of Juvenile Justice (FDJJ). This provides criminal history, basic demographics, sanctions information (incarceration facility, facility details, etc.), and recidivism data. This data set, which covers all youths released from a FDJJ program between July 2006 and July 2011, has been matched to administrative records from the Florida Department of Law Enforcement and the Florida Department of Corrections. These sources provide information on adult arrests, convictions and sanctions. The final source of data is the Community Positive Achievement Change Tool (C-PACT), which provides information on criminal history, success in school, use of free time, work experience, relationships, family life, drug and alcohol use, mental health, attitudes/behaviors, aggression and social/emotional skills. The C-PACT is used to evaluate the youths' risk-to-reoffend as well as to evaluate the effectiveness of FDJJ programs. C-PACT information is gathered from criminal history records, from a structured interview with the youths, and from the youths' Juvenile Probation Officers (JPO). The JPOs have first hand knowledge of the youth's home, family, school and social lives.

I have the data from the C-PACT examination taken immediately prior to incarceration for all youths from July 2007 to July 2011. Among the youths that I have C-PACT information for, 80% of them have undergone the full examination and 20% have only a shortened version, which includes key questions in most domains but does not include the psychological measures.[3] I also have two full-length post-release C-PACT examinations for a subset of the

[2]Any such experiment would require careful monitoring to ensure that results are as anticipated (Carrell et al., 2013).

[3]The full C-PACT examination is shown in the online appendix, section A. The questions that are highlighted are also found on the shorter evaluation.

youths in residential treatment facilities: one taken immediately after release from the facility and one taken an average of eight months after release.[4] Since the C-PACT examination is administered in the home town by a counselor who is unlikely to have knowledge of the peer cohort (due to the large degree of geographical dispersion in residential facilities) any error in the C-PACT evaluation should be orthogonal to peer characteristics.

The two main outputs of the C-PACT examination are the 'life-risk' score and the 'criminal history' score, which are used by the FDJJ to evaluate the youth's risk-to-reoffend.[5][6] The life-risk score, sometimes referred to in this paper as simply the risk score, describes the extent to which a teenager's life circumstances - home, school and social life - put him/her at risk of criminal behavior. Summary statistics for some of the indicators used to build the life-risk score are shown in Panel A of Table 1: incarcerated parents, foster care placement, physical and sexual abuse, gang affiliation, whether or not s/he is a high school dropout, etc. The criminal history score measures the youth's criminal experience. Summary statistics for some of the indicators used to build the criminal history score are shown in Panel B of Table 1: age at first offense, number and type of previous crimes, and previous periods of incarceration.[7]

I also use a variety of attitude and behavioral measures from the full C-PACT examination as outcome variables. Examples are shown in Panel C of Table 1. These characteristics are summarized by an index of crime-oriented attitudes and non-cognitive factors as well as an index of healthy (crime-protective) non-cognitive factors.[8] For more detailed analysis, I have compiled the questions from the C-PACT into five categories and built measures for each. The 'aggression' category includes all questions having to do with anger, physical or verbal violence, or belief in the necessity of aggression. The 'impulsivity' measure includes all questions having to do with the inability to control impulses or emotional urges. Questions that measure lack of respect for law, the rules of society, authority figures, or the property of others, as well as belief in the necessity of physical or verbal aggression, are captured by the 'anti-societal attitudes' measure. All questions having to do with the ability to get along well

[4] I have the first post-release C-PACT for 68% of the youths in my sample; this group consists of youths who were released onto probation and whose probation officers correctly followed the guidelines of administering a C-PACT immediately upon release. I have the second post-release C-PACT for 57% of the youths. This group consists of adolescents who stayed on probation for at least three months and had a C-PACT administered some time after the third month.

[5] Florida uses the term 'social history' instead of life-risk, but as this phrase is not very meaningful I will use the term life-risk, or simply risk, in this paper.

[6] These two numerical variables are used to classify youths into four categorical risk-to-reoffend levels.

[7] The questions that are used to build the life-risk score are indicated by an arrow on the left hand side of the C-PACT document in the online appendix, section A. Domain 1 of the C-PACT evaluation shows all questions used to build the criminal history score.

[8] These scores are the average of the standardized 'risky' and 'protective' domain scores for the three C-PACT domains that cover psychological factors.

with or relate to others are used to calculate the 'social skills' measure. Finally, 'consideration for the future' measures optimism about the future, consideration of the consequences of actions, as well as the ability to plan and enact goals.[9]

These five measures are predictive of future crime - the first three positively and the latter two negatively (results not shown). As seen in Table 2 the life-risk score, the criminal history score and both indices of crime-oriented non-cognitive factors (the one taken before entry and the one immediately after exit) also predict future crime, even after controlling for demographic and other criminal history variables. The criminal history score is the strongest predictor of recidivism, although the non-cognitive measures taken immediately after release from the facility also predict recidivism very well.

Figure 1a shows that the life-risk score is only weakly correlated with the criminal history score. This is within a sample of youths who have generally lived very difficult lives and have a lot of criminal experience; the weak correlation among this unique sample should not be interpreted as evidence that crime is only weakly correlated with the level of life-difficulties. In Figure 1b, however, we see that the life-risk score is strongly correlated with the index of crime-oriented non-cognitive factors. This is consistent with research by psychologists which shows that exposure to trauma or high levels of stress during childhood can lead to emotional and behavioral problems throughout life (Evans and Fuller-Rowell, 2013; Felitti et al., 1998).

All of the descriptive statistics shown above are from the sample of youths in residential correctional facilities. Residential facilities have a median occupancy of 40 youths and a median length of stay of eight months. Genders are separated and facilities are either low, moderate, high or maximum security. Youths generally do not leave the facility during the length of their sentence. While residential facilities are the primary focus of this analysis, I conduct supplemental analysis of youths in day treatment facilities. Similar to an alternative school for delinquent youth, the adolescent lives at home but spends weekdays at the day treatment facility, attending classes and participating in rehabilitative programming. The median occupancy is 38, the median length of stay is five months, and genders are mixed within a facility. Descriptive statistics for the two facility types are shown in Table 3. Both facility types are mostly male and African American, with an average age of around 16.5. Youths in day treatment facilities have a slightly lower criminal history and life-risk score than those in residential facilities. Recidivism rates are quite high: 50/60% are re-arrested and 15/21% are re-incarcerated within a year of release for those in day treatment/residential facilities respectively.

I drop all facilities with less than 50 observations. These were facilities for which I

[9]More details on the construction of these measures is provided in the online appendix, section C.

only had data spanning a short time period; as such the peer groups within them were highly correlated with one another. This left 160 facilities and 12,695 adolescents in the the residential sample and 29 facilities holding 3306 adolescents in the day treatment sample.[10]

2.2 Facility assignment

If the judge assigns the youth to residential placement, an employee of the FDJJ known as a commitment manager decides which facility to send the youth to. Some facilities specialize in substance abuse, mental health, or sex offense treatment; if the youth has need for such a specialized program this will influence the choice of placement.[11] Co-offenders are intentionally separated, and the commitment manager must choose facilities within a certain security-level category – low, moderate, high or maximum – which has been determined by the judge.

While the research design is similar to those which leverage cohort level variation after controlling for school and grade fixed effects (Hoxby, 2000; Carrell and Hoekstra, 2010) it differs in several ways that bolster identification. First of all, the youth's parents do not have the power to transfer him to another correctional facility in the case of a particularly undesirable peer cohort. The youth is under the jurisdiction of the juvenile justice system. Second of all, the administrator in charge of placing the youth within a residential facility is unlikely to take the peer cohort into account when deciding where to place the youth. Counties sends youths to a median of 62 different residential correctional facilities over the course of a year, and facilities receive youths from a median of 16 different counties. This high degree of geographical dispersion implies that commitment managers have very little knowledge of any particular cohort at any particular time. Third, the residential facilities are so geographically dispersed that local trends in crime will only marginally impact the type of youths present at any one residential facility. Finally, I have access to very detailed information about each facility. If a facility changes management or changes a program name, I include an extra fixed effect in order to account for any potential changes in the distribution of youths assigned to this facility.[12]

If the judge assigns a youth to a day treatment facilities, she will be placed in the one nearest to her home. Day treatment facilities are entirely local implying that local trends in

[10]Additional data details can be found in the online appendix, section D.

[11]The C-PACT score does not affect the administration of services within a particular facility/program, which mitigates concerns about a crowding-out effect.

[12]I have also conducted diagnostic tests of each facility to examine whether the distribution of youths remain stable over time. In a few circumstances there was evidence of non-random sorting that was not explained by a change in management or program name. I either dropped these facilities from the analysis or added a fixed effect to absorb the change in youth's distribution. This only occurred in a tiny fraction of my sample and does not affect results.

crime will affect the distribution of youths in a facility in a manner not likely to be absorbed by the statewide time trends. For this reason I include facility-specific linear time trends in the main specification for day treatment facilities.

2.3 Peer measures

With the exception of Section 5 which explores heterogeneity in functional form, the independent variables are a weighted average of the peer characteristics where the weights are the number of days that the two youths' sentences overlap. The calculation of the peer measures is defined in Equation 1, where $peerSCORE_i$ is the independent variable for person i, $d_{i,k}$ is the number of days that person i's sentence overlaps with person k within the same facility, and $ownSCORE_k$ is the score/trait of person k. For example, a person's peer risk score would be a weighted average of their peers life-risk score.

$$peerSCORE_i = \frac{\sum_{k \neq i} d_{i,k} * ownSCORE_k}{\sum_{k \neq i} d_{i,k}} \tag{1}$$

I use the full four years for which I have C-PACT data (July 1, 2007 – June 30, 2011) in calculating most of the peer measures. However the peer measures are calculated with error for youths released towards the beginning or end of my sample period since the data doesn't cover members of their peer group who were released before the sample period begins or after it ends. In order to minimize bias due to measurement error I drop the youths who were released in the outer ends of my sample period and perform analysis only on those who were released after April 1, 2008 and before September 31, 2010.

Figures 1c and 1d show the standardized distribution of the peer risk/criminal history score in residential facilities, shown in the left-most box-plots labeled 'Total Variation'. The bottom and top lines of the boxes show the 25th and 75th percentile of the distributions, the outer lines show the adjacent values. These raw peer measures are correlated with observable characteristics of the youths, as seen in the second-to-left box-plot labeled 'Endogenous Part of Total Variation' which shows the fitted values of a regression of peer risk/criminal history on a variety of covariates.[13] Figures 1c and 1d also show the distribution of the identifying variation – the residuals from a regression of peer risk/criminal history score on facility and quarter-by-year of release fixed effects. This is shown in the second-to-right box-plots labeled 'Identifying Variation'. While the fixed effects remove a part of the total variation, a sizable

[13]The covariates included are the same as those in the main specification as described in Section 4 with the exception of the person's own risk score when the dependent variable is peer risk and the person's own criminal history score when the dependent variable is peer criminal history. These scores are omitted due to the mechanical negative correlation described in Section 4.

amount still remains; the standard deviation of the identifying variation is approximately one third of the standard deviation of the total peer risk score and one quarter of the standard deviation of peer criminal history. Importantly, however, the fixed effects remove the part of the peer measure variation that is correlated with the observable characteristics of the youth. This can be seen in the right-most box-plot labeled 'Endogenous Part of Identifying Variation', which shows the fitted values from a regression of the identifying variation on the covariates.

3 Identification

Juvenile correctional facilities in Florida are quite small - the median occupancy is 40 youths - implying that the average characteristics of a facility's cohort will fluctuate as youths enter, serve their sentences, and leave. Peer influence is identified using this 'natural' variation that occurs across cohorts within the same correctional facility. The identifying assumption is that the variation in peer characteristics which is left over after both facility and time fixed effects have been accounted for is as-good-as-random. In day treatment facilities, which are more susceptible to local trends, the identifying assumption also conditions on facility-specific linear time controls. While I present a qualitiative argument for exogeneity in section 2.2, this section shows empirical evidence that variation in key independent variables appears as-good-as-random after the fixed effects and time trends have been accounted for.[14]

3.1 Testing for non-random clustering within facilities

I begin by regressing the life-risk and the criminal history score on a set of peer group dummies. Since peer groups are overlapping, the peer group dummies are based on those present in a particular facility within a six month window. This regression is shown in Equation 2 where $ownSCORE_{ijt}$ is the life-risk/criminal history score of person i in facility j who was released in quarter-by-year t, ω_{js} is a dummy that equals one if the person was in facility j during the six month window s, λ_j are facility fixed effects, η_t are quarter-by-year of release fixed effects, and $\mathbb{I}[day](time_{ijt} * \mu_j)$ are facility specific time trends, used only in the regressions involving the sample of youths in day treatment facilities.

$$ownSCORE_{ijt} = \alpha + \omega_{js} + \lambda_j + \eta_t + \mathbb{I}[day](time_{ijt} * \mu_j) + \epsilon_{ijt} \tag{2}$$

[14]I focus on in this section on the life risk and criminal history score of the peer cohort. In work not shown here I have also tested for, yet found no evidence of, non-random variation in other peer characteristics.

I run both an F test for joint significance at the 10% level on ω and a false discovery rate procedure to see if any individual coefficient in the ω vector is rejected when the p-values are adjusted to control for a 10% false discovery rate (Benjamini and Hochberg, 1995). ω contains 528 dummies in the residential sample and 146 dummies in the day treatment sample. Results are shown in Panel A of Table 4, where the first row shows the F statistic for a test of joint significance on ω, the second row shows the probability of seeing a statistic greater than or equal to F under the null, and the third row shows the smallest FDR adjusted p value. The first and second column shows the results for the risk and criminal history score in the residential treatment facilities and the final two columns show the results for day treatment facilities.

As can be seen in Panel A of Table 4, there is no evidence of non-random clustering in risk or criminal history score within either residential or day treatment facilities. The p values for the F test on joint significance range from .11 to .24 and the false discovery rate adjusted p values for individual significance are all greater than .20.

3.2 Testing for correlations between peer type and own type using a 'split-sample cluster bootstrap'

In most circumstances random assignment implies that the dependent variable that has been randomized is orthogonal to the covariates. When the dependent variable is a peer trait this is not the case; random assignment implies an expected negative mechanical correlation between a person's own trait and the average trait of her peers. Guryan et al. (2009) explain intuitively that this has to do with selection without replacement. In any finite population sample, a person with a higher type will be selecting peers from an urn that has a lower average type than a low type person. In a two person sample, the person with the higher type is guaranteed to have the lower type peer and vice versa. Angrist (2013) provides an analytic demonstration of this mechanical negative correlation. As is clear from both the intuitive explanation and the analytical proof, the problem stems from the fact that a person's own type shows up in the average peer type of his group.

This negative mechanical correlation complicates identification tests as it can obscure positive correlation between own and peer types.[15] Guryan et al. (2009) propose a method to correct for this negative mechanical correlation by controlling for the leave-one-out average score in the 'urn' (pool of potential peers) from which the peers are chosen; in my simulations I find that this technique is very low power (Stevenson, 2015).[16] I present here a 'split-sample

[15] Negative mechanical correlation only disappears in large samples if the number of fixed effects are relatively small (Wang, 2009).

[16] An accounting identity shows that when urn fixed effects are included, as would normally be the case

cluster bootstrap' method of testing for correlations between own and peer type that does not exhibit negative mechanical correlation under random assignment.[17]. The procedure is similar to a cluster bootstrap method in which instead of single observations being chosen randomly with replacement the entire cluster (correctional facility) is chosen randomly with replacement (Cameron et al., 2007). However in each iteration the sample is split randomly in two halves: the 'peer' half and the 'analysis' half. The peer scores for the youths in the 'analysis' sample are calculated using only youths in the 'peer' sample. Then these peer scores are regressed on own scores as shown in Equation 3, where $peerSCORE_{ijt}$ refers to the average peer risk/criminal history score of person i in facility j who is released in time period t, calculated only using observations in the 'peer' half of the split sample, and all other variables are as described above. This technique eliminates the negative mechanical correlation and yields an estimate of β that will be centered at zero under random assignment. It works because none of the scores for the 'analysis' group are used in the calculation of the peer scores for their peers; in other words, each youth's score will show up only on one side of the equation. The technique is summarized as follows:

$$peerSCORE_{ijt} = \alpha + ownSCORE_{ijt}\beta + \lambda_j + \eta_t + \mathbb{I}[day](time_{ijt} * \mu_j) + \epsilon_{ijt} \tag{3}$$

1. Randomly split the sample into two groups, a 'peer' group and an 'analysis' group.

2. Calculate the average peer score *only for* the people in the 'analysis' group, and *only using* people in the 'peer' group. Thus for all youths in the 'analysis group', their own score will not show up in the peer score of others.

3. Sample the clusters (facilities) with replacement using a 'pairs cluster bootstrap' technique (Cameron et al., 2007).

4. Run the regression specified in Equation 3, and collect the coefficient $\hat{\beta}$.

5. Repeat steps 1-4 10,000 times.

---

when cohorts are chosen from different populations, this technique is identical to controlling for ones own score scaled by the number of potential peers in the urn. When urn sizes are similar, as in the simulation example presented in Guryan et al. (2009), this added control is almost perfectly correlated with the one's own score. In this situation, controlling for the leave-one-out urn mean eliminates not only the negative mechanical correlation but also greatly reduces the ability to detect potentially problematic positive correlations. While these points are either directly or indirectly made in Guryan et al. (2009) the problems with this technique are worth emphasizing.

[17] A more detailed discussion of this technique, optimized for peer groups which are not overlapping, can be found in Stevenson (2015)

Panel B from Table 4 shows the results of the split sample test. Row one shows the average $\hat{\beta}$ after 10,000 repetitions, and row two shows the standard deviation of $\hat{\beta}$. A t-test for these estimated parameters would fail to reject the null at any level smaller than 0.54. However, the magnitude of the potential correlation is at least as important is its statistical significance: if it is not possible to rule out a sizable positive correlation between the peer and own scores then the conditional exogeneity assumption is on weaker ground. Row three of Panel B shows the upper bound for a 95% confidence interval on β using the estimated coefficients, row four shows the standard deviation for the risk/criminal history scores and row five shows the standard deviation for the peer risk/peer criminal history scores.[18] The estimates of the upper bound are quite conservative, since the split-sample technique increases the variance in the estimator, but nonetheless they are tiny in magnitude compared to the magnitudes of the independent and dependent variables. The upper bound of the 95% confidence interval implies that a one standard deviation increase in a person's own score predicts at most a 0.01-0.03 standard deviation increase in average score of their peer cohort, making a confound between own and peer types an unlikely concern.

In results not shown here I conduct tests for correlations between the peer risk/criminal history scores and the covariates. Among the 32 separate tests (eight covariates, two settings, and two peer measures) there are two that are significant at the 10% level, no more than what is expected under random assignment. Furthermore, a joint test of significance in which the peer score is the left hand side variable and all the covariates are on the right easily fails to reject in all four specifications.

4 Mechanisms of peer influence and results

The predictions that are tested in this section vary along three dimensions: the *type* of peers that are expected to be most influential under a certain mechanism, the *setting* in which the mechanism is most likely to show up, and the post-release *outcomes* that correspond with the mechanism. These predictions, as well as the strength of the evidence supporting them, are summarized in Table 5.

The generalized specification for the regressions conducted in this section is shown in Equation 4. Y_{ijt} stands for a variety of post-release outcomes for person i in facility j who is release in quarter-by-year t. $peerSCORE_{ijt}$ stands for a variety of weighted average pre-entry peer characteristics as defined in Equation 1. Covariates X_{ijt} include race, gender, ethnicity,

[18] I did not standardize these since each split of the data set would result in a slightly different scaling, making it difficult to compare the β coefficients.

age at release, risk score, criminal history score, total number of prior felony charges, total number of prior misdemeanor charges, an index that weighs the seriousness of prior charges and dummies for the category of most serious charge (violent felony, property felony, other felony and misdemeanor). The covariates also include interactions between the facility fixed effects and the risk and criminal history score. Finally, I condition on fixed effects for facility (λ_j), fixed effects for quarter-by-year of release (η_t), and, in the day treatment sample, facility-specific linear time trends ($\mathbb{I}[day](time_{ijt} * \mu_j)$).

$$\begin{aligned} Y_{ijt} = \alpha + peerSCORE_{ijt}\gamma_0 + X_{ijt}\beta_0 + \lambda_j + \eta_t + \\ \mathbb{I}[day](time_{ijt} * \mu_j) + \epsilon_{ijt} \end{aligned} \quad (4)$$

[19]

4.1 Social Contagion of Attitudes, Behaviors and Non-cognitive Traits

Mechanism Description: This mechanism captures the influence that peers have over non-cognitive factors related to crime: aggressive behavioral habits, lack of impulse-control, risk preferences, anti-societal attitudes, and so forth. Peers may influence social norms, potentially de-stigmatizing aggressive, non-cooperative, or illegal behavior (Posner, 1997). Peers may also influence identity formation, altering the psychological payoff of different behaviors according to how much they deviate from or support that identity (Akerlof and Kranton, 2000). For example if a youth adopts an 'outlaw' or a 'gangster' self-identity, he is likely to increase behaviors which conform to that image. Peers can also influence beliefs: if a youth has low expectations for success in the 'straight world' then the opportunity cost of crime is lower. Finally, peers may influence the development of social or emotional skills directly through social learning (Ellison and Fudenberg, 1995). The development of anger management, for example, rests on a variety of tricks such as counting to ten before responding, or consciously re-formulating the interpretation of the other person's action. If peers possess few of these skills there are less opportunities by which to learn them.

While all of these channels can directly influence preferences for crime, they may also indirectly influence the net financial returns to crime by influencing the non-cognitive skill set.

[19] Angrist (2013) discusses a variety of econometric issues in peer effects research, two of which are particularly relevant to this specification. The first, mechanical negative correlation, is addressed by the inclusion of one's own score as well as the peer score as a regressor. In my simulations this was effective at eliminating negative bias. The second has to do with measurement error, or other reasons why an IV estimate of the impact of own score on outcomes might be greater than OLS. However this is not a serious concern in situations where peer groups are random or quasi-random, since the peer score is not a good instrument for own score.

Different non-cognitive skills may be important for different types of tasks. Characteristics such as aggression are not generally considered skills in the legal economy but may prove beneficial in a line of work where the ability to instill fear in others helps with contract enforcement (drug dealing, for example). Alternatively, peer influence could decrease the opportunity cost of crime by shifting non-cognitive factors away from those valued in the legal economy: persistence, social skills, optimism, etc. (Bowles et al., 2001; Heckman and Rubinstein, 2001; Nagin and Pogarsky, 2004; Heckman et al., 2006, 2013; Blanden et al., 2007; Moffitt et al., 2011).

The unifying characteristic of this mechanism is that peers are influential not because they offer some practical advantage (knowledge, skill, joint productivity, etc.) but because they alter the youth's psychological landscape.

Predictions: It is a folk wisdom trope that teenagers are very susceptible to peer influences. Nonetheless, shifting rooted habits of behavior and thought is likely to take a fairly intensive social experience. For this reason the social contagion channel of influence is likely to show up more strongly in residential facilities, where teenagers spend 24 hours a day together isolated from the rest of the world, than day treatment facilities, where they spend the schooldays together but then return to their friends and families. The peers that are most likely to be influential in this mechanism are those with high levels of crime-oriented non-cognitive factors such as aggression or anti-societal attitudes. Exposure to such peers should result not only in an increase in crime, but also an increase in crime-oriented non-cognitive traits.

Results: Table 6 presents the results of a variety of tests related to the social contagion mechanism of peer influence. The tests all are based on the residential facility sample of youths; all independent variables have been standardized for ease of interpretation. I begin by examining the influence of peers with a high life-risk score on recidivism outcomes. As seen in Figure 1a and b, the life-risk score is strongly correlated with crime-oriented non-cognitive factors, but only weakly correlated with criminal experience or skill. Influence of peers with a high life-risk score on post-release crime would thus be more consistent with the social contagion mechanism than with the other mechanisms.

As seen in Panel A of Table 6, exposure to peers with a high life-risk score leads to a large increase in post-release crime. The specification for this regression is shown in Equation 4, where Y_{ijt} stands for five different recidivism dummies indicating whether or not person i is arrested, arrested for a felony, convicted, convicted of a felony, or placed back in a juvenile or adult prison within a year after release.[20] The independent variable ($peerSCORE_{ijt}$) is

[20]The various recidivism measures are correlated but still measure different things; the correlation coefficients range from 0.4 to 0.7.

the weighted average life-risk score of the peer cohort as defined in Equation 1, and the fixed effects and covariates are as described in the introduction to this section.

The magnitude of this effect is comparable in predictive power to an additional 2 or 3 prior felonies on the criminal record. Transferring an adolescent from a peer cohort with an average life-risk score to a cohort whose average life-risk is one standard deviation higher predicts a 16% increase over the mean in the likelihood of being re-incarcerated within a year of release. With p values ranging from 0.003 to 0.021 the effects are also highly significant.[21]

Table 7 shows that the impact peers with a high life-risk score have on recidivism are extremely stable. The regression in the first column includes only fixed effects for facility and quarter-by-year of release. The second through fifth columns add demographics, criminal history variables, the life-risk score, and both the life-risk score and the criminal history score interacted with facility fixed effects. Column 5 includes all the same covariates and fixed effects of the main specification described in the introduction. Column 6 adds extra controls for linear time trends interacted at the facility level and Column 7 includes peer criminal history, peer age, and the percent of peers who are African American or Hispanic. (These peer traits are built using the same weighted average method described in Equation 1.) The coefficient magnitude varies only trivially across specifications, increasing slightly as controls for own risk score are added (consistent with mechanical negative correlation) and barely budging with the addition of controls for the criminal history score of the peer cohort and facility-specific linear time trends.

The life-risk score is a composite measure that captures many aspects of the youth. Panel B of Table 6 divides the peer risk score into its component parts in an attempt to see which factors are most influential on post-release crime. Columns 1 through 3 look at the impact of peers with gang affiliation, peers from unstable homes, and peers who have experienced abuse or trauma on the likelihood of being convicted of a felony within one year of release.[22] Peers with gang affiliation are not influential in this setting, supporting the hypothesis that this effect is not driven by network formation. The peers that are most influential on future crime are those from difficult or dangerous homes.

Psychologists have found that difficulties in early childhood can have long lasting effects

[21]Some of the econometric quirks in peer effects research that were outlined in Angrist (2013) suggest that parametrically estimated standard errors may not be correct. To check this I conduct a permutation test to verify that the standard errors presented in Panel A of Table 6 are not too small. The permutation test involves randomly shuffling the entry dates of youths within their facility, generating randomly assigned counter-factual peer cohorts which should have no effect on post-release outcomes. With each randomly generated permutation I run the five regressions used in Panel A of Table 6. I derive two-tailed p-values from this non-parametric distribution of γ_0 under the null hypothesis. These non-parametrically p-values turn out to be quite similar to those generated parametrically.

[22]The final two components of the peer risk score, peers with drug/alcohol problems and peers who are experiencing problems at school, were not found to be statistically significant predictors of future crime.

on emotional development and executive functioning. A broadly accepted theory for why this occurs is ‘allostatic load’ – the idea that if the brain’s stress-management system is repeatedly overworked it begins to break down from the strain (McEwen, 1998). Studies have found that even after controlling for socioeconomic factors, the amount of traumatic experiences a person had as a child is a very strong predictor of risky behaviors, mental health issues, poor executive functioning and other undesirable outcomes (Evans and Fuller-Rowell, 2013; Felitti et al., 1998). Research also suggests that behavioral problems stemming from childhood trauma can be contagious: Carrell and Hoekstra (2010) find that classroom exposure to peers who have experienced domestic abuse leads to an increase in misbehavior. While their setting differs from the one analyzed here, the story is similar: difficult life circumstances leads to emotional and behavioral problems with externalities on surrounding youth.

In Column 4 of Panel B I test directly whether exposure to peers with crime-oriented non-cognitive traits impacts recidivism. I only have data on pre-entry non-cognitive factors for a subset of the youth, meaning that this independent variable will have a fair amount of measurement error. Nonetheless the coefficient is large in magnitude and significant at the 5% level.

The large majority of youths in residential facilities come from home towns that are far enough apart to prohibit much physical interaction after release, making the network formation mechanism an unlikely explanation for the results shown in Panel A. However approximately 10% of the each youth’s cohort comes from the same hometown area. Column 5 tests whether the effects examined in Panel A are being caused by hometown peers; the independent variable is the weighted average life-risk score of peers within a 20 mile radius of the youth’s home zip code. As can be seen, this effect is tiny in magnitude and not statistically significant.

Panels C and D of Table 6 show how exposure to peers with a high life-risk score affects post-release non-cognitive outcomes. In addition to the covariates and fixed effects discussed in the introduction to this section, I include a fully saturated set of controls for the pre-entry non-cognitive factors. I also include both fixed effects and specific time trends for the DJJ unit which conducts the examination in order to absorb some of the incidental variation in grading styles. These additional controls add power but do not qualitatively affect results. The dependent variables in Panel C were taken immediately after release from the facility and those in Panel D were taken approximately eight months after release. All dependent variables have been standardized.

Exposure to high risk peers while incarcerated leads to an increase in aggression, impulsivity, anti-societal attitudes and an index score that captures a range of crime-oriented non-cognitive factors. Once again, the magnitude of the effects are considerable: a one standard deviation increase in the peer risk score leads to a 0.11-0.16 standard deviation increase in the various crime-oriented non-cognitive measures and a 0.13 standard deviation decrease in an index score of healthy non-cognitive traits.[23] Furthermore, as shown in Panel D, these effects persist at least eight months post-release. While the magnitudes of the effects are slightly smaller, it is still statistically significant in four of the five specifications despite the smaller sample size. Panel D also shows that the non-cognitive effects captured eight months after release do not appear to be caused by the subset of youths from the same hometown area, suggesting that the influence was exerted predominantly during the incarceration period.

In sum, the evidence presented in this section shows that exposure to peers who have grown up in difficult circumstances, and who show high levels of aggresssion and anti-societal attitudes leads to an increase in recidivism, an increase in crime-oriented behaviors and attitudes, and a decrease in healthy non-cognitive functioning. This phenomenon shows up in an environment of intense social exposure, and one in which peers are unlikely to have much physical interaction with each other after release.[24] This effect is extremely stable to different specifications, including those that control for the criminal experience of the cohort. Considered comprehensively, the evidence is strongly consistent with the social contagion mechanism and would be difficult to reconcile with either skill transfer or persistent network formation.

4.2 Persistent Network Formation

Mechanism description: One of the defining characteristics of the illegal economy is its covert nature. The conventional methods by which economic information is transmitted - advertising, job posts, credit scores - are unadvisable for those wishing to avoid detection by the law. Information in the illegal economy travels largely through interpersonal connections, implying that the returns to an expansion of the criminal network may be sizable. A criminal network provides benefits similar to those studied in labor economics (Jackson, 2011): transmitting information about opportunities (which house is easy to break into), reducing information asymmetries (who can be trusted not to 'snitch'), connecting buyers with sellers (distributors of weapons or drugs), providing informal insurance (an advance

[23] In results not shown, I also test directly whether exposure to peers with high levels of aggression lead to increased aggression, etc. The results are positive and statistically significant for aggression, anti-societal attitudes, impulsivity, social skills, the index of crime-oriented traits and the index of healthy traits.

[24] I do not find evidence supportive of the social contagion mechanism in the day treatment centers. The coefficient on peer risk is not statistically significant and varies in sign. Furthermore there is no evidence that peers affect non-cognitive outcomes in this setting.

of marijuana after one's supply has been seized), accelerating technology uptake (advanced forms of identity fraud), etc. The lack of a formal venue for contract enforcement suggests an additional role for networks in the illegal economy: contract enforcement via the threat of retaliation.

The unifying characteristic of this mechanism is that peers influence crime through interactions that occur after release from the correctional facility.[25]

Predictions: The network formation mechanism posits that incarceration fosters the formation of valuable connections much like business school, or an elite prep school. However, for the connections to be valuable after release, they need to be relevant to a particular criminal market. Most adolescent crime is highly local – few teenagers have criminal operations that are sophisticated enough to span across the state – thus the connections that will be most influential in post-release crime are those with others from the same locality. Network formation is expected to be much more relevant in day treatment facilities, which unite youths from the same local area, than residential facilities, where the average distance between home towns of two peers in the same facility is 100 miles. If the network formation mechanism showed up at all in residential facilities it would likely operate through the small group of peers who return to neighboring home towns after release.

The peers that would be important in the network formation hypothesis are peers who are likely to commit crime after release, or peers who have lots of criminal connections themselves. Since past criminal behavior is the best predictor of future criminal behavior, peers with a high criminal history score are likely to be influential in the network formation mechanism. Peers with gang affiliation are likely to be influential as well, as exposure to these well connected peers could lead to a broader range of connections after release. In addition to post-release crime, the network formation mechanism predicts an increase in gang affiliation and network based crimes such as drug dealing.

Results: Table 8 summarizes the evidence that supports the network formation mechanism of peer influence. All independent variables have been standardized, and all regressions take the basic format outlined in the introduction to this section.

In residential facilities, the overall criminal history of the peer group has no effect on future crime, despite the fact that criminal history is the best predictor of recidivism (see Column 7 of Table 7 and Panel A of Table 9). In day treatment facilities, where the youths are all local, the results are quite different. In Panel A of Table 8 we see that day-treatment exposure to peers with a high criminal history score has a large impact on future crime:

[25] One paper that makes a strong case for being able to pinpoint a mechanism of peer influence is that by Drago and Galbiati (2012). They leverage a policy that affects *post-release* criminal activity of the peer cohort to show that networks formed during the incarceration period persist after release and affect recidivism.

a one standard deviation increase in the average criminal history score of the peer cohort predicts a 38% increase over the mean in the likelihood of being convicted of a felony within a year of release. However the day treatment sample is roughly one fourth the size of the residential sample and the peer effects are much noisier. The coefficient on peer criminal history is statistically significant at the 5% level in only two of the five specifications, and statistically significant at the 10% level in another two.

In residential facilities, exposure to peers with gang affiliation has no effect on future crime. In day treatment facilities these well-connected peers are more influential. Panel B shows that day-treatment exposure to peers with gang affiliation leads to an increase in felony conviction and re-incarceration. Although the coefficients are positive in all specifications, the results are only statistically significant in these latter two measures of recidivism.

Returning to the residential setting, I find exposure to peers with a high criminal history score has an effect, but *only if* those peers come from the same hometown area. This is shown in Panel C of Table 8, where the independent variable is the weighted average criminal history score of peers who live within 20 miles of one's own hometown.

In Panel D I test whether exposure to peers with a high criminal history score affects gang membership or drug dealing, which is typically a crime controlled by gangs. While the estimated effects are positive, they are only statistically significant when the facility-specific time trends are removed. In results not shown here I also test whether exposure to gang members increases gang membership or gang-related crimes. The effects are positive but not statistically significant.

Finally Column 5 of Panel D provides a brief robustness check. The regression shown in this column is identical to that shown in Column 4 of Panel A except that all covariates have been removed. The magnitude of the coefficient is quite stable.

In sum, the evidence presented in this section suggests that exposure to peers with lots of criminal experience or criminal connections leads to an increase in post-release crime and gang affiliation, but only if these peers live close enough to one another to enable physical interaction after release. This effect is strong in day treatment centers, but also occurs in a more limited way in residential facilities. While the evidence presented in this section is noisier than that presented in the social contagion mechanism, it is internally consistent and best explained by the formation of new criminal networks that persist after release.

4.3 Criminal Skill Transfer

The theoretical model that underpins this mechanism of peer influence is the classic Mincer model, in which labor market returns increase as skills are acquired (Mincer, 1974).

Whereas the skills necessary for employment in the legal economy are generally acquired through school or through job training programs, no such equivalent exists in the illegal economy. Skills are likely gained through experience or through transfer between individuals. Thus the criminal skill transfer theory is based on the idea that correctional facilities can act as a 'school of crime': an informal clearing house for illicit know-how. With lots of time and little to do, inmates may share expertise in areas like disabling car alarms, fencing stolen computers or manufacturing methamphetamine.

Predictions: The criminal skill transfer hypothesis predicts that exposure to criminally experienced peers would lead to an increase in crime.[26] Since the mechanism also requires unsupervised time within the facility during which the youths can teach each other skills, it is more likely in the residential facilities than in day treatment facilities. (Most of the youth's time in the day treatment facilities is under active supervision, either during the school day, or in the after-school programming.) The criminal skill transfer mechanism also predicts that exposure to peers with experience in a particular crime category leads to an increase in the likelihood of recidivating with that particular crime.

Results: Table 9 shows the results of tests related to the skill transfer mechanism. The independent variable in Panel A is the weighted average criminal history score of the entire peer cohort; the sample is from the residential facilities. Since the independent variable captures the criminal experience of the peer cohort – a proxy for skill – a positive and statistically significant impact on recidivism would provide support for the skill transfer mechanism. However the coefficients are small in magnitude, vary in sign, and are not statistically significant.

Panels C-E show how exposure to peers with experience in a particular skill-intensive crime category affects the likelihood of committing that crime after release Following Bayer et al. (2009), I test for heterogenous effects in Panels B and D under the hypothesis that those with previous experience in that particular crime category may be more susceptible. Since the specification in these panels does not involve the life-risk or criminal history score (which are only available in four of the five years for which I have data) I omit these covariates in order to take advantage of the larger sample size.

The outcomes are dummies measuring whether or not the youth is charged or adjudicated guilty for a robbery, felony drug offense, burglary, auto theft, or grand larceny within a year of release. The independent variables vary according the outcome variables, but are all placed in the same row so as to conserve space in the table. For example, when the outcome

[26]Skilled youths are also less likely to be caught implying that the criminal history score will be an imperfect measure of a youth's true experience. However the probability of getting caught increases with the number of crimes committed and in this paper I assume that the criminal history score is an imperfect but still useful measure of the youth's criminal experience.

variable is robbery, the independent variable is the fraction of peers who have committed robbery in the past. In Panels B and D these are interacted with dummies indicating that the youth has also committed robbery in the past (first row) or a dummy indicating the youth has not committed robbery (second row). Following Bayer et al. (2009) I control for previous experience in all the offense categories listed, as well as offense interacted with facility fixed effects. These extra controls, however, make no qualitative difference to the findings.

Of the 20 tests conducted in these four panels, only one is significant at the 5% level, and the sign on that coefficient is negative. These results replicate the findings in Damm and Gorinas (2013) and are partially similar to those in Bayer et al. (2009).[27] Of these five crime categories, Bayer et al. (2009) find effects only for drug offenses and burglary – and only if the youth has previous experience in that particular crime category. (Similar to Bayer et al. (2009), I find results for the crime categories of sex-offense and assault. As these are not skilled crimes I do not include the results in this section.)

In sum, I do not find evidence to support the skill transfer mechanism. While the lack of supportive evidence may simply be due to low power or measurement error, it could be argued that many criminal skills involve the type of manual techniques that would be difficult to learn in an institutional setting. As anyone who has tried to repair a car or assemble a piece of furniture by reading a manual knows, physical tasks can challenging without physical instruction. Picking locks, disabling alarms or motion detectors, hot-wiring cars – like other physical skills, these techniques are likely best learnt "on the job" as opposed to in whispered conversations in the prison yard.

5 Heterogeneity and policy implications in the social contagion mechanism

In Section 4 we saw that placing teenagers with a cohort which, on average, had a higher degree of behavioral problems and life difficulties lead to an increase in post-release crime. In this section I use a more flexible functional form for peer influence in order to determine whether negative externalities can be diminished by sorting the youths differently within incarceration facilities. To explore the impact of peers with different life-risk levels I divide the risk score into bins of two, as seen in Figure 2a. I aggregate the tails of the distribution so that each bin contains at least 10% of the sample. Then for each youth I build six different

[27] Damm and Gorinas (2013) only finds results for drug crimes if the peer group is narrowly defined to be those of similar age or ethnicity.

peer measures: the fraction of peers in each of the six bins. This is expressed in Equation 5, where $frac_i^b$ is the fraction of i's peers (weighted by days) with risk score in bin b.

$$frac_i^b = \frac{\sum_{k \neq i} d_{i,k} * \mathbb{I}[k \in b]}{\sum_{k \neq i} d_{i,k}} \tag{5}$$

Dropping the third bin (so that the fraction of peers with a risk score between six and seven will serve as the reference level) I regress felony conviction on the five peer risk-level variables. The set of covariates and fixed effects are the same as in all the residential facility regressions as specified in Equation 4. I present the results of this regression in graphical form in Figure 2b. The independent variables are represented on the x axis and the coefficient magnitudes are represented on the y axis. The dropped variable is set equal to zero on this graph to serve as a reference point. The shape of the points represent statistical significance, as described in the legend. Each point (each coefficient) represents the marginal impact of having more peers in that particular bin, conditional on the number of peers you have in each other bin. Thus if you experience a 10% increase in the number of the highest risk peers in your cohort, your likelihood of being convicted of a felony increases by 2.5 percentage points. Since an average of 27% of all youths released from incarceration are convicted of felony within a year, the 10% increase in very high risk peers translates into a 9% increase over the mean in felony convictions. While only results for felony conviction are discussed in this section, the functional form for the other recidivism variables are qualitatively quite similar.

This section is presented graphically so that the focus is on looking for descriptive patterns, not on seeking statistical significance of each individual coefficient. Each coefficient represents a relatively low power test, as it is much more difficult to distinguish the impact of peers in two adjacent risk-level bins than it is to identify a general relationship between the risk level of peers and future crime, like the linear-in-means test does.

Figure 2b shows several interesting results. First of all, the coefficients are monotonic and reasonably linear, implying that the effect operates throughout the spectrum and does not solely come from the impact of one type of peer. Second of all, there appears to be an outlier effect, where exposure to particularly high risk peers significantly increases recidivism, even controlling for the fraction of peers in other bins. However these results aggregate both genders, obscuring differences in functional form.

Figure 2c examines how the impact of peer risk on felony conviction differs by gender. The coefficients shown are the result of two regressions, one for boys and one for girls, of felony conviction on the five peer risk level variables. For boys, the fraction of peers with low to moderate risk levels appears to have very little influence, while the fraction of high risk peers increases recidivism markedly. For girls the opposite is true: the fraction of peers with moderate to high risk levels has little influence, but the fraction of peers with very low risk levels decreases recidivism markedly.[28]

It is not immediately clear why girls respond more to positive peer influence than boys do, however this finding corresponds with similar findings in the literature. The benefits of early childhood education – often attributed to the learning of positive non-cognitive skills from teachers and mentors – accrue much more for girls than for boys (Anderson, 2008; Heckman et al., 2013). Kling et al. (2007) finds that moving to higher SES neighborhoods – presumably implying a lower-risk peer group – improves mental health and risky behavior outcomes for girls, while having generally negative effects for boys.

Figures 2d and e show how peers at different risk levels affect felony conviction for youths who are low, medium and high risk themselves. Figure 2d shows the results for boys and Figure 2e shows results for girls. Each graph shows the coefficients from a single regression of felony conviction on the six peer risk level variables interacted with dummies for being in the bottom, middle and top third of the risk distribution. The impact of peers with risk score between six and seven for the low risk youths are dropped from the equations. Small sample sizes make it difficult to draw definitive conclusions about the functional form across the three risk groups.

The non-linearities in the impact of peers with different risk levels suggests that outcomes for low and medium risk boys might be improved by separating them from the boys with higher levels of aggression and emotional disturbance. I divide the male sample into three groups - low risk boys whose life-risk score is five and under, medium risk boys whose score is between six and nine, and high risk boys whose life-risk score is ten or above. For each group of boys I estimate the impact of replacing a low risk peer with a medium or high risk peer, a medium risk peer with a low or high risk peer, or a high risk peer with a low or medium risk peer on their likelihood of being convicted of a felony or re-incarcerated within a year. Using these estimated coefficients, I do a back of the envelope calculation for the expected recidivism rates of each of these three groups under an alternative sorting regime in which the youths are entirely segregated into groups of other similarly ranked youths (low, medium or high risk). I then compare these outcomes to those observed in the data, where the average fraction of low risk boys in a facility is 0.33 (with a standard deviation of 0.10), the average fraction of medium risk boys is 0.45 (with a standard deviation of 0.07) and the average fraction of high risk boys is 0.22 (with a standard deviation of 0.09).

[28] A test for the equality of coefficients rejects that the influence of very low risk peers is the same across genders at the 5% level, and rejects that the influence of very high risk peers is the same across genders at the 10% level.

I estimate that outcomes are improved for the low and medium risk boys under this alternative method of organizing, while outcomes are worsened for the high risk boys. The net effects, however, are positive. When boys are sorted so that their peers are all in the same risk category, my calculations predict a 16% decrease in the number of boys who are convicted of a felony and an 18% decrease in the number of boys who are re-incarcerated within a year.

An average of 6030 boys are released each year from a residential correctional program in the state of Florida during the years for which I have data (July 2006 to July 2011). 15% of these boys are back in a juvenile incarceration facility within a year of release and 7% are in an adult prison or jail. The average cost of incarcerating an adolescent is $27,000 in a juvenile facility and $15,415 in an adult facility.[29] The estimated yearly savings from fully sorting the youths into low, medium and high risk cohorts would exceed $5.6 million in terms of incarceration costs alone, not counting the social gains of reduced crime. While sorting the youths according to their risk level would lead to a definite welfare loss for high risk boys, this could be potentially offset by increasing their access to evidence-based rehabilitative programming, paid for out of savings from the lower incarceration rates.

These numbers are of course only back-of-the-envelope estimates.[30] Furthermore the youths may respond endogenously to the different peer distributions in ways that are difficult to predict (Carrell et al., 2013). Nonetheless the estimated benefits of separating the lower risk boys from those with higher levels of aggression/anti-societal attitudes are substantial, and as long as the youths are carefully monitored to make sure that the re-sorting is having the expected positive effects, it seems an advisable experiment to undertake.

6 Conclusion

This paper presents evidence that social interactions had while incarcerated can have a strong impact on the propensity for crime after release. I find that exposure to peers from unstable homes and who have emotional/behavioral problems leads to a sizable increase in criminal behavior after release as well as an increase in crime-oriented attitudes and behav-

[29]These cost estimates are generated assuming an eight month stay, which is the average in juvenile facilities. As the length of incarceration in adult facilities is generally longer this is likely to be an underestimate. The yearly cost of incarcerating an adult inmate is $20,553 (Henrichson and Delaney, 2012) and the daily cost of incarceration in a juvenile facility is $97.92 in a non-secure facility and $141.62 in a secure facility (SPLC, 2010).

[30]It could be argued that these numbers are an underestimate since they do not include the social gains of averted crime or the benefits to the youths themselves. On the other hand the practical constraints of implementing the new sorting regime may result in some welfare loss: separating the lower risk boys from the higher risk boys may require placing them in a facility that is too far from home for their parents to be able to visit, for example.

ioral habits. This effect is most consistent with a social contagion mechanism: adolescents are influenced by the attitudes, beliefs, values and behaviors of their peers in a way that alters their criminal propensity. Non-linearities in peer influence suggest that net outcomes can be improved by sorting the youths to minimize negative peer influence. Back-of-the-envelope calculations estimate that separating lower risk boys from higher risk boys would lead to an 18% decrease in the one year re-incarceration rate, with $5.6 million in savings simply from reduced incarceration costs alone.

Although the focus of the paper is on peer influence in juvenile incarceration facilities, I conduct a supplementary analysis on peer influence in day treatment facilities (alternative schools for delinquent youth). Day treatment facilities differ from the residential facilities in a manner that is important for the persistent network formation mechanism: they unite youths from the same local area. In residential facilities I find that exposure to peers with lots of criminal experience leads to an increase in crime after release, but only if those peers live close enough to one another that they are likely to interact after release. This finding is confirmed and expanded upon in day treatment setting: exposure to peers with lots of criminal experience as well as peers with criminal connections (gang affiliation) leads to an increase in crime after release, as well as an increase in gang participation. The finding that this type of peer is only influential when they come from the same local area is consistent with the network formation mechanism.

While the evidence presented in this paper speaks most directly to the question of how to sort youths in order to minimize negative peer influence, it also raises questions about a variety of other important policy questions. Are peer influences in juvenile incarceration damaging enough that home arrest might be a preferable alternative to a group-based form of sanctions? What are the consequences in terms of exposing adolescents to criminally experienced adults when they are placed in an adult jail or prison? If peers are so influential, is there a way to harness this in a positive manner - through mentoring, support groups, or programs that target social cohesion?

Incarceration is a heavily-used tool in the United States criminal justice system. In 2011 there were an estimated 61,423 youths in a juvenile incarceration facility, and 2,240,600 people (both adults and adolescents tried as adults) in state prison, federal prison, or jail (Sickmund et al., 2013; Glaze and Herberman, 2013). Despite the scale of this intervention, much is still unknown about its effects on future outcomes. The evidence presented in this paper underline the fact that youths who are incarcerated are not simply sitting on the shelf: the incarceration period is actively formative. *Who* a teenager is locked up with influences his attitudes towards society, beliefs around law and justice, aggressive behaviors, emotional and social skills. *Who* a teenager is locked up with influences how likely he is to be locked up again, a year after release. While the primary rationale for incarceration is to lower crime through deterrence or incapacitation, the primary experience of incarceration is one of being confined in a small physical space with a particular group of people: criminally experienced and often emotionally troubled. Understanding how this experience affects adolescents is a step towards helping them from getting caught in the revolving door between crime and incarceration.

References

Aizer, Anna and Joseph J. Jr. Doyle, "Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-Assigned Judges," *National Bureau of Economic Research Working Paper*, 2013.

Akerlof, George A. and Rachel E. Kranton, "Economics and Identity," *The Quarterly Journal of Economics*, 2000, *115* (3), pp. 715–753.

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

Angrist, Joshua, "The Perils of Peer Effects," Working Paper 19774, National Bureau of Economic Research December 2013.

Bayer, Patrick, Randy Hjalmarsson, and David Pozen, "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections," *Quarterly Journal of Economics*, 2009, *124*, 105–147.

Benjamini, Yoav and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, *57* (1), 289–300.

Billings, Stephen B., David J. Deming, and Jonah Rockoff, "School Segregation, Educational Attainment and Crime: Evidence from the end of busing in Charlotte-Mecklenburg," *The Quarterly Journal of Economics*, 2013.

Blanden, Jo, Paul Gregg, and Lindsey Macmillan, "Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education*," *The Economic Journal*, 2007, *117* (519), C43–C60.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne, "The Determinants of Earnings: A Behavioral Approach," *Journal of Economic Literature*, 2001, *39* (4), 1137–1176.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," Working Paper 344, National Bureau of Economic Research September 2007.

Carrell, Scott E. and Mark L. Hoekstra, "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," *American Economic Journal: Applied Economics*, 2010, *2* (1), 211–28.

_, Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 2013, *81* (3), 855–882.

Chen, M. Keith and Jesse M. Shapiro, "Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-based Approach," *American Law and Economics Review*, 2007, *9* (1), 1–29.

Damm, Anna Piil and Cedric Gorinas, "Deal Drugs Once, Deal Drugs Twice: Peer Effects on Recidivism from Prisons," Technical Report, Working Paper December 2013.

Drago, Francesco and Roberto Galbiati, "Indirect Effects of a Policy Altering Criminal Behavior: Evidence from the Italian Prison Experiment," *American Economic Journal: Applied Economics*, 2012, *4* (2), 199–218.

Duckworth, Angela L. and Christopher Peterson, "Grit: Perseverance and Passion for Long Term Goals," *Journal of Personality and Social Psychology*, 2007, *92* (6), 1087–1101.

Ellison, Glenn and Drew Fudenberg, "Word-of-Mouth Communication and Social Learning," *The Quarterly Journal of Economics*, 1995, *110* (1), pp. 93–125.

Evans, Gary W. and Thomas E. Fuller-Rowell, "Childhood poverty, chronic stress, and young adult working memory: the protective role of self-regulatory capacity," *Developmental Science*, 2013, *16* (5), 688–696.

Felitti, VJ1, RF Anda, D Nordenberg, Williamson DF, AM Spitz, V Edwards, MP Koss, and JS Marks, "Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study.," *American Journal of Preventative Medecine*, 1998, *14* (4).

Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman, "Crime and Social Interactions," NBER Working Papers 5026, National Bureau of Economic Research, Inc February 1995.

Glaze, Lauren E. and Erinn J. Herberman, "Correctional Populations in the United States 2012," Technical Report, Bureau of Justice Statistics December 2013.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo, "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 34–68.

Hazel, Neal, "Cross-National Comparison of Juvenile Crime Rates," Technical Report, Youth Justice Board 2008.

Heckman, James J. and Yona Rubinstein, "The Importance of Noncognitive Skills: Lessons from the GED Testing Program," *The American Economic Review*, 2001, *91* (2), pp. 145–149.

__ , Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, July 2006, *24* (3), 411–482.

Heckman, James, Rodrigo Pinto, and Peter Savelyev, "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 2013, *103* (6), 2052–86.

Henrichson, Christian and Ruth Delaney, "The Price of Prisons: What Incarceration Costs Taxpayers," Technical Report, Vera Institute of Justice January 2012.

Hoxby, Caroline, "Peer Effects in the Classroom: Learning from Gender and Race Variation," Working Paper 7867, National Bureau of Economic Research August 2000.

Jackson, Matthew O., "Chapter 12 - An Overview of Social Networks and Economic Applications," in Alberto Bisin Jess Benhabib and Matthew O. Jackson, eds., , Vol. 1 of *Handbook of Social Economics*, North-Holland, 2011, pp. 511 – 585.

Jacob, Brian A. and Lars Lefgren, "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime," *American Economic Review*, 2003, *93* (5), 1560–1577.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Landenberger, NanaA. and MarkW. Lipsey, "The positive effects of cognitivebehavioral programs for offenders: A meta-analysis of factors associated with effective treatment," *Journal of Experimental Criminology*, 2005, *1* (4), 451–476.

Lerman, A.E., *The Modern Prison Paradox: Politics, Punishment, and Social Community*, Cambridge University Press, 2013.

Manski, Charles F, "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, July 1993, *60* (3), 531–42.

McEwen, Bruce, "Stress, Adaptation and Disease: Allostasis and Allostatic Load," *Annals New York Academy of Science*, 1998.

Mendel, Richard A., "No Place for Kids: The Case for Reducing Juvenile Incarceration," Technical Report, The Annie E. Casey Foundation, Baltimore, MD 2011.

Mincer, Jacob A., *Schooling, Experience, and Earnings*, Columbia University Press, 1974.

Moffitt, Terrie E., Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J. Hancox, HonaLee Harrington, Renate Houts, Richie Poulton, Brent W. Roberts, Stephen Ross, Malcolm R. Sears andW. Murray Thomson, and Avshalom Caspi, "A gradient of childhood self-control predicts health, wealth, and public safety," *PNAS*, 2011.

Nagin, Daniel S. and Greg Pogarsky, "Time and Punishment: Delayed Consequences and Criminal Behavior," *Journal of Quantitative Criminology*, 2004, *20* (4), 295–317.

Ouss, Aurelie, "Prison as a School of Crime: Evidence From Cell Level Interactions," Technical Report, Working Paper December 2011.

Posner, Richard A., "Social Norms and the Law: An Economic Approach," *The American Economic Review*, 1997, *87* (2), pp. 365–369.

Sickmund, M, T.J. Sladky, W Kang, and C Puzzanchera, "Easy Access to the Census of Juveniles in Residential Placement," Technical Report, Office of Juvenile Justice and Delinquency Prevention 2013.

SPLC, "Fiscal Responsibility: The Key to a Safer, Smarter, and Stronger Juvenile Justice System," Technical Report, Southern Poverty Law Center December 2010.

Stevenson, Megan, "Tests of random assignment to peers in the face of mechanical negative correlation: an evaluation of four techniques," Working Paper March 2015.

Wang, Liang Choon, "Peer Effects in the Classroom: Evidence from a Natural Experiment in Malaysia," Technical Report, Working Paper January 2009.

7 Tables and Figures

Table 1: Summary statistics on life-risk, criminal history, and non-cognitive indicators

	Male	Female	Total
Panel A: Summary Statistics for the Life-Risk Score			
Dropped out or expelled	0.11	0.13	0.11
Drinks alcohol	0.21	0.27	0.22
Does drugs	0.46	0.44	0.46
Known gang affiliate	0.14	0.10	0.14
Parents w/drug, alc. prob	0.14	0.22	0.15
Parent has been incarcerated	0.45	0.52	0.46
Family member has been incarcerated	0.60	0.67	0.61
Been in foster care/shelter	0.17	0.32	0.19
Ran away or was kicked out	0.42	0.77	0.47
Physically abused	0.16	0.36	0.19
Sexually abused	0.030	0.29	0.065
Panel B: Summary Statistics for the Criminal History Score			
First offense under 14yrs	0.78	0.76	0.77
First offense under 12yrs	0.39	0.33	0.38
Previously incarc.	0.29	0.28	0.29
MS prior- assault	0.39	0.56	0.42
MS prior- burglary	0.35	0.16	0.33
MS prior- robbery	0.14	0.037	0.13
MS prior- drug felony	0.059	0.040	0.057
MS prior- auto theft	0.043	0.080	0.048
MS prior- larceny	0.11	0.30	0.14
MS prior- murder, att. murder	0.0043	0.0029	0.0041
Total prior felony charges	6.06	2.85	5.62
Total prior misdemeanor charges	5.91	4.97	5.78
Panel C: Summary Statistics for the Non-Cognitive Scores			
Low/no hope or aspiration	0.40	0.39	0.40
Impulsive or highly impulsive	0.57	0.64	0.58
Lacks empathy for victim	0.28	0.33	0.29
Lacks respect for other's property	0.17	0.15	0.17
Low tolerance for frustration	0.20	0.36	0.22
Hostile interp. of other's actions	0.50	0.63	0.51
Believe verb. agg. often necessary	0.10	0.21	0.12
Believe phys. agg. often necessary	0.066	0.10	0.071
No/unrealistic goals	0.56	0.55	0.56
Poor situational perception	0.81	0.82	0.81
Poor social skills	0.81	0.79	0.81
Poor emotional skills	0.60	0.65	0.60

The statistic shown is the mean.

Note: This table shows summary statistics by gender on some of the C-PACT indicators used to build the life-risk, criminal history, and non-cognitive scores for 12,695 youths in residential facilities. All variables except for the final two variables in Panel B are dummies indicating whether or not a trait is observed. 'MS prior' is short for 'most serious prior offense'.

Table 2: Do the risk, criminal history, and non-cognitive scores predict recidivism?

Panel A: Not including non-cognitive scores					
	(1)	(2)	(3)	(4)	(5)
	Arrest	Felony Arrest	Convict	Felony Convict	Prison
Own Risk	0.0309****	0.0242****	0.0195****	0.0113***	0.0169****
	(0.00420)	(0.00402)	(0.00454)	(0.00373)	(0.00334)
Own CH	0.0861****	0.0907****	0.0697****	0.0561****	0.0474****
	(0.00426)	(0.00518)	(0.00475)	(0.00475)	(0.00384)
Observations	12696	12696	12696	12696	12696
R^2	0.121	0.128	0.107	0.0829	0.136
Mean dep. var.	0.60	0.44	0.43	0.27	0.21
Panel B: Including non-cognitive scores					
	Arrest	Felony Arrest	Convict	Felony Convict	Prison
Own Risk	0.0211****	0.0155***	0.0160***	0.00693	0.0144***
	(0.00541)	(0.00558)	(0.00558)	(0.00475)	(0.00467)
Own CH	0.0758****	0.0808****	0.0597****	0.0572****	0.0450****
	(0.00560)	(0.00719)	(0.00595)	(0.00620)	(0.00512)
Crim-Noncog (Entry)	0.0218***	0.0191***	0.00500	-0.000938	0.00000405
	(0.00744)	(0.00725)	(0.00727)	(0.00641)	(0.00679)
Crim-Noncog (Exit)	0.0345****	0.0275****	0.0388****	0.0262****	0.0377****
	(0.00536)	(0.00632)	(0.00525)	(0.00583)	(0.00534)
Observations	7702	7702	7702	7702	7702
R^2	0.123	0.123	0.121	0.0948	0.150
Mean dep. var.	0.64	0.47	0.47	0.29	0.25

Standard errors clustered at the facility level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows how the life-risk, criminal history, and the non-cognitive scores taken both before entry and after exit from the facility predict recidivism among youths in residential facilities. The independent variables have been standardized and the dependent variables are binary, indicating whether or not the youth is arrested/convicted/imprisoned within a year of release. The regression is linear and includes controls for demographics, facility fixed effects, and quarter-by-year of release fixed effects.

Table 3: Summary statistics on demographics, criminal history and recidivism, shown by facility type

	Residential	Day treatment	Total
Pre-determined Variables			
Age at entry	16.5	16.3	16.4
Female	0.14	0.23	0.16
African-American	0.55	0.55	0.55
Hispanic	0.11	0.15	0.12
MS prior-violent felony	0.61	0.45	0.58
MS prior-property felony	0.29	0.31	0.30
Previously incarcerated	0.29	0.22	0.27
Own risk score	7.19	5.38	6.82
Own criminal history score	12.8	8.94	12.0
Recidivism Variables			
Arrest 1yr post	0.60	0.50	0.58
Felony arrest 1yr post	0.44	0.32	0.42
Conviction 1yr post	0.43	0.34	0.41
Felony conviction 1yr post	0.27	0.18	0.25
Re-incarcerated 1yr post	0.21	0.15	0.20

The statistic shown is the mean.

Note: This table shows summary statistics by facility type. There are 12,695 observations for the residential facilities and 3,306 observations for the day treatment facilities. 'MS prior' stands for 'most serious prior offense'. The recidivism variables are dummy variables indicating whether or not the adolescent was arrested/convicted/imprisoned within one year of release.

Table 4: Identification tests: within-facility cluster test and the 'cluster bootstrap split sample test' for non-random assignment to peers

Panel A: Testing for non-random clustering of types within facilities				
	Residential-Risk	**Residential-Crim. History**	**Day-Risk**	**Day-Criminal History**
F Statistic on ω	1.04	1.08	1.13	1.15
Prob >F under null	0.2423	0.1125	0.1330	0.1082
Smallest FDR p value	0.8258	0.8821	0.4967	0.2034
Panel B: Testing for correlations between own risk/CH score and peer risk/CH				
	Residential-Risk	**Residential-Crim. History**	**Day-Risk**	**Day-Criminal History**
$\hat{\hat{\beta}}$ (Average $\hat{\beta}$: 10,000 reps)	0.0017	0.0018	0.0022	0.0016
$\hat{\hat{\sigma}}_\beta$ (Average $\hat{\sigma}_\beta$: 10,000 reps)	0.0032	0.0028	0.0036	0.0031
Upper Bound, 95% C.I.	0.0081	0.0074	0.0094	0.0078
S.D. Own Score	1	1	1	1
S.D. Peer Score	0.38	0.46	0.31	0.40

Note: Panel A summarizes the results of two different tests for non-random clustering of types within a facility. The first test is an F test for joint significance on a set of peer group dummies. The dependent variable is the risk/criminal history score. The second test is an FDR adjusted test for individual significance on the same set of dummy variables. Panel B summarizes the results for the 'cluster bootstrap split-sample' test of correlation between own risk/criminal history score and the risk/criminal history score of the peers. Descriptions for both techniques can be found in Section 3.

Table 5: A summary of tests and evidence for the three mechanisms

Social Contagion

Setting	Peer Characteristic	Outcome	Evidence?
Residential	↑ Life-risk score	↑ Crime	Strong
Residential	↑ Life-risk score	↑ Crime-oriented noncogs	Strong
Residential	↑ Life-risk score	↓ Protective noncogs	Strong
Residential	↑ Crime-oriented noncogs	↑ Crime	Strong
Residential	↑ Crime-oriented noncogs	↑ Crime-oriented noncogs	Strong(Not Shown)
Residential	↑ Crime-oriented noncogs	↓ Protective noncogs	Strong(Not Shown)
Residential	↑ Abusive/unstable homes	↑ Crime	Strong

Network Formation

Setting	Peer Characteristic	Outcome	Evidence?
Day	↑ CH score	↑ Crime	Strong
Day	↑ Gang affiliation	↑ Crime	Medium
Day	↑ CH score	↑ Gang affiliation	Medium
Day	↑ Gang affiliation	↑ Gang affiliation	Weak(Not Shown)
Day	↑ CH score	↑ Gang-related crimes	Medium
Residential	↑ CH score, hometown peers	↑ Crime	Medium

Skill Transfer

Setting	Peer Characteristic	Outcome	Evidence?
Residential	↑ CH score	↑ Crime	None
Residential	↑ Specific skilled crime	↑ Specific skilled crime	None
Day	↑ Specific skilled crime	↑ Specific skilled crime	None

Note: This table summarizes the predictions of the various mechanisms as well as the degree to which a prediction is supported by the evidence. For example the first row in the social contagion panel should be interpreted as follows: "The social contagion mechanism predicts that an increase in the life-risk score of the peer group in the residential setting should lead to an increase in crime after release. A test of this prediction provides strong confirmatory evidence." For more details about how these tests relate to the different mechanisms see Section 4. 'CH' stands for 'criminal history'.

Table 6: Social Contagion

Panel A: Residential Facilities					
	(1) Arrest	(2) Felony Arrest	(3) Convict	(4) Felony Convict	(5) Prison
Peer Risk	0.0377***	0.0405***	0.0312**	0.0349***	0.0344****
	(0.0139)	(0.0137)	(0.0132)	(0.0129)	(0.0100)
Observations	12695	12695	12695	12695	12695
Mean dep. var.	0.60	0.44	0.43	0.27	0.21

Panel B: Residential Facilities: Alternative Specifications					
	Felony Convict	Felony Convict	Felony Convict	Felony Convict	Felony Convict
Peer Gang	0.00927				
	(0.0101)				
Peer Unstable Home		0.0253*			
		(0.0132)			
Peer Abuse/Trauma			0.0293**		
			(0.0133)		
Peer Crim. Noncog				0.0257**	
				(0.0126)	
Peer Risk, Hometown					0.00335
					(0.00541)
Observations	11251	12695	12695	12695	12695
Mean dep. var.	0.27	0.27	0.27	0.27	0.27

Panel C: Residential Facilities: Non-cognitive Outcomes Post-Release					
	Aggr-ession	Anti-Societal Attitudes	Impul-sivity	Crime-Oriented Traits	Healthy Traits
Peer Risk	0.108**	0.160***	0.150***	0.158***	-0.131**
	(0.0493)	(0.0552)	(0.0526)	(0.0528)	(0.0527)
Observations	7035	7035	7035	7035	7035

Panel D: Residential Facilities: Non-cognitive Outcomes 8 Months Post					
	Aggr-ession	Anti-Societal Attitudes	Impul-sivity	Crime-Oriented Traits	Healthy Traits
Peer Risk	0.0984*	0.128**	0.0907	0.128**	-0.123*
	(0.0533)	(0.0565)	(0.0571)	(0.0533)	(0.0621)
Peer Risk, Hometown	-0.0151	0.0151	0.00748	0.000250	0.00234
	(0.0153)	(0.0155)	(0.0196)	(0.0158)	(0.0172)
Observations	5310	5310	5310	5310	5310

Standard errors, clustered at facility level, are shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All regressions are linear and include the full set of covariates and fixed effects shown in Column 5 of Table 7. Panels A/B: The independent variables have been standardized and the dependent variable is binary, indicating whether or not the youth is arrested/convicted/imprisoned within a year of release. Panels C/D: Both the independent and the dependent variables have been standardized. These regressions also include fully saturated controls for pre-entry non-cognitive measures and linear time trends for the DJJ unit that administers the examination.

Table 7: Social Contagion: Stability of effect

	(1) Felony Convict.	(2) Felony Convict.	(3) Felony Convict.	(4) Felony Convict.	(5) Felony Convict.	(6) Felony Convict.	(7) Felony Convict.
Peer Risk	0.0354***	0.0382***	0.0355***	0.0384***	0.0349***	0.0353**	0.0355**
	(0.0132)	(0.0131)	(0.0127)	(0.0129)	(0.0129)	(0.0161)	(0.0167)
Peer CH							-0.0198
							(0.0190)
Perc. Afr.-Am.							-0.102
							(0.0903)
Av. Age							0.0528*
							(0.0311)
Perc. Hisp.							-0.0936
							(0.147)
Facility FE	X	X	X	X	X	X	X
Date FE	X	X	X	X	X	X	X
Demographics		X	X	X	X	X	X
Prior Crimes			X	X	X	X	X
Risk Score				X	X	X	X
Risk x FacFE					X	X	X
CH x FacFE					X	X	X
LinearTT x FacFE						X	X
Observations	12695	12695	12695	12695	12695	12695	12695
R^2	0.0536	0.0709	0.0869	0.0876	0.102	0.110	0.111
Mean dep. var.	0.27	0.27	0.27	0.27	0.27	0.27	0.27

Standard errors, clustered at facility level, are shown in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Peer Risk has been standardized and the dependent variable is a dummy variable indicating whether or not the youth has been convicted of a felony within a year after release. Covariates include facility fixed effects, quarter-by-year of release fixed effects, demographics, criminal history variables and score, risk score, both criminal history and risk score interacted with facility fixed effects, a linear time trend interacted with facility fixed effects, peer criminal history, percent of African American peers, average age of peers and percent of Hispanic peers.

Table 8: Network Formation

Panel A: Day Facilities: Criminal History Score of Peers					
	(1)	(2)	(3)	(4)	(5)
	Arrest	Felony Arrest	Convict	Felony Convict	Prison
Peer CH	0.0646*	0.0427	0.0966***	0.0734**	0.0611*
	(0.0355)	(0.0419)	(0.0322)	(0.0277)	(0.0355)
Observations	3306	3306	3306	3306	3306
Mean dep. var.	0.51	0.32	0.34	0.18	0.15
Panel B: Day Facilities: Peers With Gang Affiliation					
	Arrest	Felony Arrest	Convict	Felony Convict	Prison
Peer Gang	0.0193	0.0267	0.0286	0.0398**	0.0423**
	(0.0221)	(0.0166)	(0.0221)	(0.0181)	(0.0158)
Observations	3306	3306	3306	3306	3306
Mean dep. var.	0.51	0.32	0.34	0.18	0.15
Panel C: Residential Facilities: Peers from Hometown Area					
	Arrest	Felony Arrest	Convict	Felony Convict	Prison
Peer CH, Hometown	0.0130**	0.0136**	0.00877	0.00670	-0.00585
	(0.00649)	(0.00628)	(0.00640)	(0.00501)	(0.00480)
Observations	11251	11251	11251	11251	11251
Mean dep. var.	0.60	0.44	0.43	0.27	0.21
Panel D: Day Facilities: Gang Related Outcomes/Robustness Check					
	Gang	Gang	Drug Off.	Drug Off.	Felony Convict
Peer CH	0.0183**	0.0140	0.0353**	0.0133	0.0610*
	(0.00933)	(0.0109)	(0.0147)	(0.0230)	(0.0286)
Facility FE	X	X	X	X	X
Time FE	X	X	X	X	X
LinearTT x FacFE		X		X	X
Covariates	X	X	X	X	
Observations	2928	1504	1504	2928	3306
Mean dep. var.	0.11	0.11	0.05	0.05	0.18

Standard errors, clustered at facility level, are shown in parenthesis. $*p < 0.10$, $**\ p < 0.05$, $***p < 0.01$

Note: Panels A/B/C: The independent variables have been standardized and the dependent variable is binary, indicating whether or not the youth is arrested/convicted/imprisoned within a year of release. The phrase 'Peer CH, Hometown' refers to the average criminal history score of youths whose zip codes are within 20 miles of one's own zip code. Panel D: The outcomes here are binary and refer to whether or not the youth has joined a gang, committed a drug offense (a gang-related crime), or been convicted of a felony within a year of release. Columns 1 and 4 omit the facility-specific time trends and Column 5 omits the covariates as a robustness check. The regression are linear and, besides the three mentioned, include a full set of covariates and fixed effects as described in Section 4.

Table 9: Skill Transfer

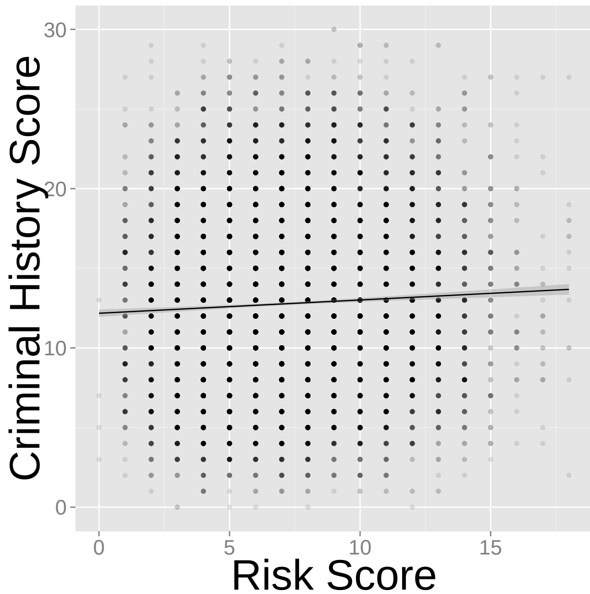
Panel A: Residential Facilities					
	(1)	(2)	(3)	(4)	(5)
	Arrest	Felony Arrest	Convict	Felony Convict	Prison
Peer CH	-0.0121	-0.00831	0.00728	-0.0110	0.00150
	(0.0147)	(0.0151)	(0.0179)	(0.0144)	(0.0164)
Observations	12695	12695	12695	12695	12695
Mean dep. var.	0.60	0.44	0.43	0.27	0.21
Panel B: Residential Facilities - crime-specific skill transfer					
	Robbery	Drug Offense	Burglary	Auto Theft	Grand Larceny
PeerOffense*Offense	0.0250	0.116	0.0836	-0.263	-0.137
	(0.0933)	(0.204)	(0.0592)	(0.241)	(0.179)
PeerOffense*NoOffense	-0.0390	0.000865	0.0355	0.0537	-0.00473
	(0.0471)	(0.0453)	(0.0395)	(0.0436)	(0.0269)
Observations	21288	21288	21288	21288	21288
Mean dep. var.	0.11	0.05	0.31	0.05	0.04
Panel C: Residential Facilities - crime-specific skill transfer					
	Robbery	Drug Offense	Burglary	Auto Theft	Grand Larceny
PeerOffense	-0.0300	0.00657	0.0502	0.0394	-0.0105
	(0.0415)	(0.0444)	(0.0317)	(0.0422)	(0.0265)
Observations	21288	21288	21288	21288	21288
Mean dep. var.	0.11	0.05	0.31	0.05	0.04
Panel D: Day Facilities - crime-specific skill transfer					
	Robbery	Drug Offense	Burglary	Auto Theft	Grand Larceny
PeerOffense*Offense	-0.406**	0.627	-0.0815	0.344	-0.171
	(0.156)	(0.720)	(0.122)	(0.385)	(0.434)
PeerOffense*NoOffense	-0.0444	0.0579	0.00889	0.0445	0.0339
	(0.0791)	(0.0781)	(0.0907)	(0.110)	(0.0702)
Observations	4442	4442	4442	4442	4442
Mean dep. var.	0.09	0.05	0.29	0.05	0.04
Panel E: Day Facilities - crime-specific skill transfer					
	Robbery	Drug Offense	Burglary	Auto Theft	Grand Larceny
PeerOffense	-0.102	0.0923	0.00281	0.0171	0.00765
	(0.0718)	(0.0920)	(0.0716)	(0.102)	(0.0717)
Observations	4442	4442	4442	4442	4442
Mean dep. var.	0.09	0.05	0.29	0.05	0.04

Standard errors, clustered at facility level, are shown in parenthesis. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

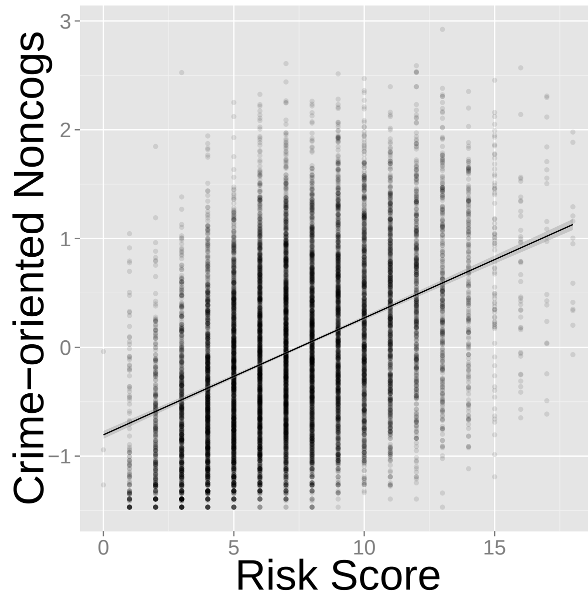
Note: Panel A: The independent variables have been standardized and the dependent variable is binary, indicating whether or not the youth is arrested/convicted/imprisoned within a year of release. Panel B-E: The outcome variables are binary and describe whether the youth commits a robbery, drug offense, burglary, auto theft, or grand larceny within a year of release. The independent variables are standardized and vary according to the outcome variable. For example, when the outcome variable is robbery, the independent variable is the fraction of peers with robbery experience. In Panels B and D this has been interacted with two dummy variables indicating whether or not the youth has robbery experience himself.

Figure 1: Correlations of traits and graphical analysis of identifying variation

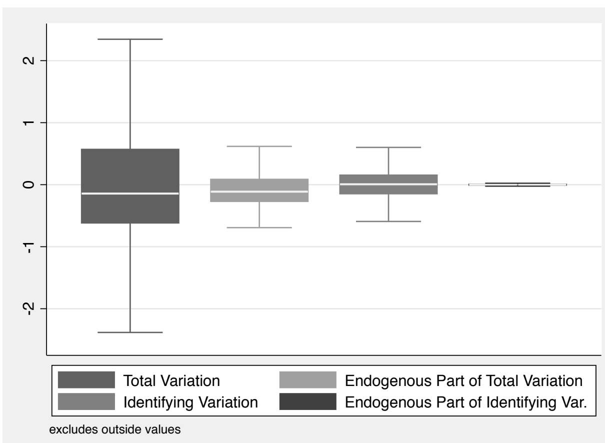
(a) Risk and Criminal History Score



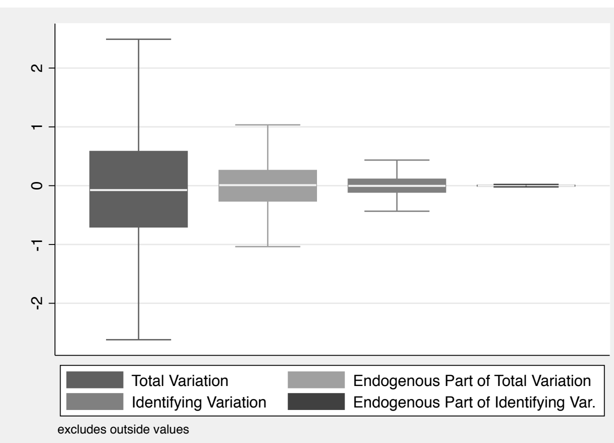
(b) Risk and Crime-Oriented Non-cognitive Index



(c) Peer Risk Score, Breakdown of Variance



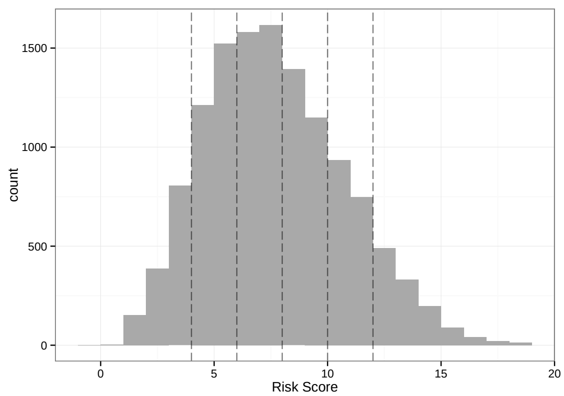
(d) Peer Criminal History Score, Breakdown of Variance



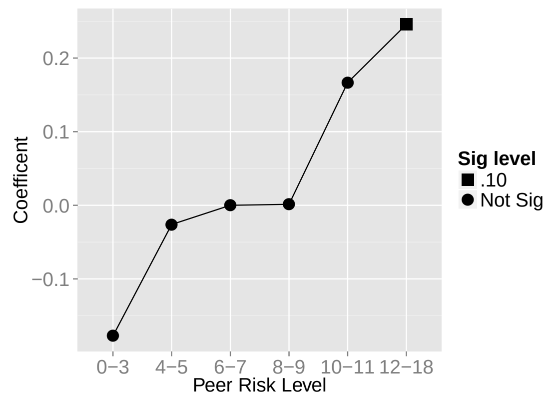
Note: The top row of graphs show the correlation of the life-risk score with the criminal history score and with the index of crime-oriented non-cognitive factors respectively. All three measures are from the examination taken immediately before entry to the facility. In the bottom row of graphs, the left-most box-plot labeled 'Total Variation' shows the distribution of the peer risk (or criminal history) score, standardized. The bottom and top lines of the box show the 25th and 75th percentile of the distribution, the outer lines show the adjacent values. The box-plot labeled 'Endogenous Part of Total Variation' (second to left) show the part of 'Total Variation' that is explained by the covariates. The box-plot labeled 'Identifying Variation' (second to right) shows the residuals of a regression of 'Total Variation' (peer risk or criminal history score) on facility and quarter-by-year of release fixed effects. The box-plot labeled 'Endogenous Part of Identifying Variation' (shown right-most) shows the part of 'Identifying Variation' that is explained by the covariates.

Figure 2: Functional Form and Heterogeneity of the Social Contagion Effect

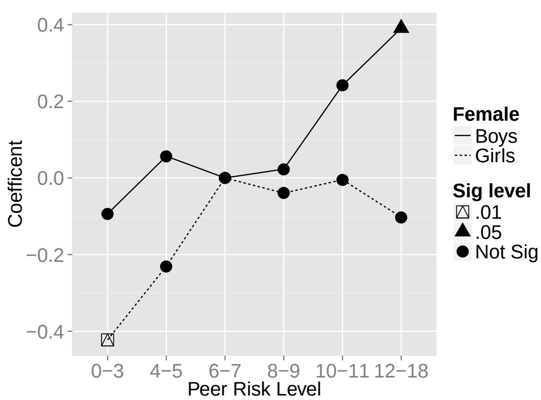
(a) Distribution of Risk Score Divided into Bins



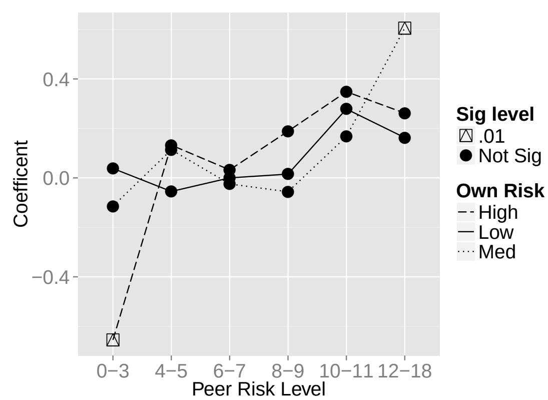
(b) Felony Conviction



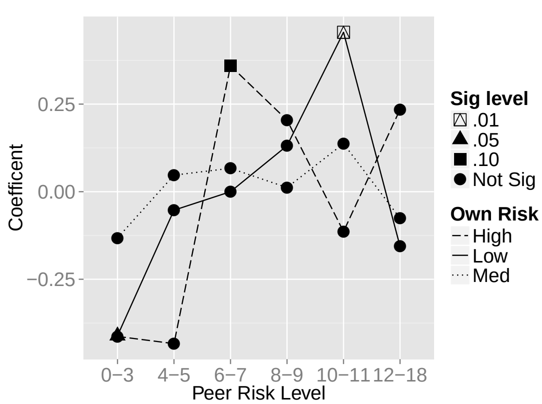
(c) Felony Conviction by Gender



(d) Felony Conviction by Own Risk, Boys



(e) Felony Conviction by Own Risk, Girls



Note: Figure 2a shows the distribution of the risk score; the vertical dotted lines delineate six bins. Figures 2b-e show the coefficients from a regression of felony conviction on the fraction of peers in each of the six bins. Figure 2b shows the full sample, Figure 2c shows results by gender and Figures d and e show results by own risk score. The fraction of peers with risk score between 6 and 7 has been dropped from the regression and is represented by a 0 in the figure. Felony conviction is a dummy variable and the crime-oriented non-cognitive index has been standardized. Each regression includes the full set of covariates and fixed effects as described in Section 4.

Article 3: Introductory Note

I present a novel method of testing for random assignment to peer groups – a 'split sample test' – and evaluate it using simulations alongside three other commonly used tests: a permutation test, a test of joint significance of peer group dummies, and a test which controls for the mean type in the urn of potential peers. All four tests are designed to address finite-sample mechanical negative correlation between peers. While the first three tests have correct size and similar power against an alternative hypothesis of positive correlation, the split sample test is the only one which provides an estimate of the magnitude of correlation between peers. The test which controls for the mean type in the urn of potential peers has low power in all simulations and incorrect size in some.

Tests of random assignment to peers in the face of mechanical negative correlation: an evaluation of four techniques

1 Introduction

Efforts to show that peers causally influence outcomes have been complicated by the fact that peer groups are often formed on the basis of pre-determined characteristics. To avoid the confound between causal effects and mere correlation, researchers seek settings in which variation in peer exposure is at least arguably random. Empirical tests which support the randomness claim are an important bolster to the causal argument. However these tests present a technical challenge due to mechanical negative correlation between characteristics of peers within a group. Unless treated appropriately, mechanical negative correlation can 'cancel out' a positive correlation making the researcher less likely to reject a false null hypothesis of random assignment to peers.

I present here a novel test of random assignment to peers which involves splitting the sample to break the mechanical negative correlation (Stevenson, 2014). I evaluate it alongside three other tests which, despite being commonly used in the literature, have not been systematically evaluated in this context. The first two were proposed by Guryan et al. (2009) and consist of a test which controls for the average score in the urn of potential peers and a permutation test. The third was proposed by Wang (2009) and consists of an F-test of joint significance of peer group dummies.

The split-sample test is the only one of the four which generates an estimate of the magnitude of correlation between peers; using this and the 95% confidence interval the researcher is able to place an upper bound on the magnitude of selection into peer groups. This may be particularly useful in circumstances where the random assignment may not have been implemented correctly (Chetty et al., 2011), or which rely on 'natural' sources of randomness (Hoxby, 2000; Stevenson, 2014).

Both the split-sample test, the permutation test and the F-test all have approximately correct rejection rates under the null in a variety of simulated scenarios. These three tests have similar power when the alternative hypothesis is positive correlation between peers, however the F-test of joint significance is not designed to reject when peers are negatively correlated. The test which controls for the mean score of the urn of potential peers has low power: using the exact same simulation parameters with which this technique was first demonstrated, the power to detect positive correlation between peers is approximately 1/9 that of the other three techniques evaluated.[1] Furthermore it has incorrect size in several simulated settings.

2 Tests for random assignment to peer groups

A commonly used specification for evaluating peer influence is to regress an outcome variable Y_i on the leave-me-out peer average of some pre-determined characteristic $\overline{X}_{-i}$ along with controls. In order to eliminate concerns about selection bias, many papers try to show that the leave-me-out peer average $\overline{X}_{-i}$ is orthogonal to X_i. However, in finite samples, $\overline{X}_{-i}$ is expected to be negatively correlated with X_i even if peer groups have been randomly assigned. This can make it difficult to determine whether or not variation in peer exposure is truly random.

Consider the regression shown in Equation 1, where X_{igu} is the pre-determined type/score X of person i in peer group g drawn from urn u, $\overline{X}_{-i,gu}$ is the leave-me-out average type of peer group g for person i, λ_u are urn fixed effects, and ϵ_{igu} is the error term. Mechanical negative correlation implies that the expectation of $\widehat{\beta}_{naive}$ is negative if peer groups are randomly assigned. Econometrically this is because each person's score shows up on both sides of the equation: X_{igu} is part of the construction of $\overline{X}_{-j,gu}$ and as such influences both the independent and dependent variables. While an increase in X_{igu} does not effect $\overline{X}_{-i,gu}$ in absolute terms, it leads to a relative decrease in $\overline{X}_{-i,gu}$ compared to the other independent variables in the vector.

$$X_{igu} = \alpha + \overline{X}_{-i,gu} * \beta_{naive} + \lambda_u + \epsilon_{igu} \tag{1}$$

GKN test: The issue of mechanical negative correlation between own and peer characteristics was first discussed in detail by Guryan et al. (2009). They propose a modification of Equation 1 which involves controlling for the leave-me-out average type of all of person i's potential peers in urn u ($\overline{X}_{-i,u}$). The goal of this control term is to adjust for mechanical negative correlation so that $\widehat{\beta}_{GKN}$ is centered at zero under random assignment. This test rejects at the α level if $|\frac{\widehat{\beta}_{GKN}}{\widehat{\sigma}_{GKN}}| > t(1-\frac{\alpha}{2})$.

[1]This technique has been used in a broad range of studies. See, for example, Carrell et al. (2009); Sojourner (2013); Fletcher et al. (2013); Eisenberg et al. (2014); Lu and Anderson (2015)

$$X_{igu} = \alpha + \overline{X}_{-i,gu} * \beta_{GKN} + \overline{X}_{-i,u} * \gamma + \lambda_u + \epsilon_{igu} \quad (2)$$

An accounting identity shows that the inclusion of the GKN correction term is mechanically identical to controlling for the person's own score scaled by the number of potential peers in her urn (Equation 3). When urn sizes are similar this control term is highly correlated with the dependent variable.

$$X_{igu} = \alpha + \overline{X}_{-i,gu}\beta_1 + \frac{-X_{igu}}{N_u - 1} * \gamma_2 + \widetilde{\lambda}_u + \epsilon_{igu} \quad (3)$$

Permutation test: Guryan et al. (2009) also propose using a permutation test to evaluate whether peers appear randomly assigned. This involves randomly shuffling the peer groups within the urns of potential peers and running the regression specified in Equation 1 with the randomly generated peer groups. Repeating this process many times generates an empirical distribution of $\widehat{\beta}_{naive}$ under the null. A test at the α level would reject if the estimate of $\widehat{\beta}_{naive}$ generated using the observed data is smaller than the $\frac{\alpha}{2}$ quantile or larger than the $1 - \frac{\alpha}{2}$ quantile of the empirical distribution.

F-test: A third test of random assignment was proposed by Wang (2009) and involves an F-test of joint significance of peer group dummies ω_g as shown in Equation 4.

$$X_{igu} = \alpha + \omega_g + \lambda_u + \epsilon_{igu} \quad (4)$$

Split-sample test: A final test of random assignment to peers was first used in Stevenson (2014) and involves splitting the sample to break the mechanical negative correlation. The technique begins by randomly selecting one observation from each peer group and calculating the peer score *only for* the people who have been randomly chosen and *only using* the scores of those who were not chosen. Then run the regression specified in Equation 5 where $i \in b, gu$ indicates the observation selected (b indicates the randomly chosen sub-sample) and $-i, gu$ indicates the unselected observations constituting person $i's$ peers. By splitting the sample, each person's score shows up only on one side of the equation: either as an independent variable or as part of another person's peer score.[2]

$$\overline{X}_{-i,gu} = \alpha + X_{i\in b,gu} * \beta_{split} + \lambda_u + \epsilon_{igu} \quad (5)$$

This regression does not exhibit mechanical negative correlation, so under random assignment the expectation of $\widehat{\beta}_{split}$ will be 0. Furthermore, $\widehat{\beta}_{split}$ provides an estimate of the correlation between own score and peer score which can be used to estimate the parame-

[2]Placing the peer score on the left hand side of the equation makes β_{split} more readily interpretable.

ters of a data generating process. For example, consider a simple data generating process in which $X_{ig} = \mu_g + v_{ig}$ with $E[v_{ig}|\mu_g, v_{jg}] = 0$. This captures either the case of random assignment (μ_g is constant across groups) or positive correlation (μ_g varies across groups). $\widehat{\beta}_{split}$ is a consistent estimator of $\frac{\sigma_g^2}{\sigma_g^2+\sigma_v^2}$, where σ_g^2 is the between group variance and σ_v^2 is the within-group variance.[3] See appendix for a more detailed discussion.

The efficiency of the split sample test can be improved by performing multiple iterations. In other words, the process of randomly choosing one observation per peer group and running the regression specified in Equation 5 is repeated B times for some large B. In each iteration b, the estimate $\widehat{\beta}_{split,b}$ as well as its parametrically estimated variance ($\widehat{\sigma}^2_{split,b}$) are collected and saved. Consider the following definitions:

$$\widehat{\beta}_{split\_mean} = \frac{\sum_{b=1}^{B} \widehat{\beta}_{split,b}}{B}$$

$$\widehat{\sigma}^2_{iter} = \sum_{b=1}^{B} \frac{(\widehat{\beta}_{split,b} - \widehat{\beta}_{split\_mean})^2}{B-1}$$

$$\overline{\widehat{\sigma}}^2_{split} = \sum_{b=1}^{B} \frac{\widehat{\sigma}^2_{split,b}}{B}$$

$$\widehat{\sigma}^2_{split\_mean} = \overline{\widehat{\sigma}}^2_{split} - \widehat{\sigma}^2_{iter}$$

$\widehat{\beta}_{split\_mean}$ is a more efficient estimator than $\widehat{\beta}_{split}$ since it eliminates much of the variance resulting from the re-sampling, and $\widehat{\sigma}^2_{split\_mean}$ provides a consistent estimate of the variance of $\widehat{\beta}_{split\_mean}$. The split-sample test rejects at the α level if $\left|\frac{\widehat{\beta}_{split\_mean}}{\widehat{\sigma}_{split\_mean}}\right| > t(1-\frac{\alpha}{2})$. For a more detailed discussion of the variance of $\widehat{\beta}_{split\_mean}$ see appendix.

The split sample technique allows the researcher to place an upper bound on the magnitude of correlation between peers ($\widehat{\beta}_{split\_mean} + 1.96 * \widehat{\sigma}_{split\_mean}$), providing useful information

[3]The variance of the empirical group means will generally be greater than σ_g^2 as it captures some of the individual level variation. There are alternative ways of estimating $\frac{\sigma_g^2}{\sigma_g^2+\sigma_v^2}$ but if peer groups differ in size or if the variation in peers is only random conditional on certain fixed effects these can be mathematically cumbersome.

beyond simply whether or not a test fails to reject a null hypothesis of random assignment. A small upper bound will help the researcher to bolster the causal argument, particularly in situations where the randomization process is not explicit or certain.

3 Simulations

I use simulations to evaluate the performance of these techniques in eight examples whose parameters are defined in Table 1. Examples 1, 3, 5 and 7 have peer groups randomly assigned ($\sigma_g^2 = 0$). Each even-numbered example is identical to the example which precedes it except that peers are positively correlated within groups ($\sigma_g^2 > 0$). Example 1 is the the same simulated example used by Guryan et al. (2009) to demonstrate how controlling for the mean score in the urn of potential peers corrects for mechanical negative correlation. Example 3 is designed to be a situation in which the F-test may not perform well – peer groups contain only two individuals and the characteristic X is not normally distributed. Example 5 is designed to mimic a research design in which classes are randomly assigned within a particular school and grade. The school-by-grade constitutes an urn and there are few peer groups per urn. In Example 7 there are ample peer groups per urn, but only a few urns.

Each simulated sample contains U urns with peer groups of size S selected from each urn. Each urn has G_u peer groups per urn, where G_u is chosen at random and with replacement from the set P with the condition that they are not all identical.[4] The data generating process is as described above: $X_{ig} = \mu_g + v_{ig}$, where $\mu_g \sim f_g$ with variance σ_g^2 and $v_{ig} \sim f_v$ with variance σ_v^2.

[4]This condition is required for the GKN method.

Table 1: A summary of different simulations

Ex. #	Name	U	S	P	f_g	f_v	$\frac{\sigma_g^2}{\sigma_g^2+\sigma_v^2}$
1	'GKN example'	100	3	{13,14,15,16,17}	0	$N(0,1)$	0
2	'GKN example'	100	3	{13,14,15,16,17}	$N(0,0.03)$	$N(0,1)$	.0291
3	'Bernoulli Pairs'	25	2	{15,20,25}	0	$Bern(P(1)=.2)$	0
4	'Bernoulli Pairs'	25	2	{15,20,25}	P(0)=.5 P(.25)=.5	$Bern(P(1)=.2)$	.0857
5	'School'	40	18	{2,3,4}	0	$N(0,1)$	0
6	'School'	40	18	{2,3,4}	$N(0,0.03)$	$N(0,1)$	0.0291
7	'Few urns'	3	18	{40,60,80}	0	$N(0,1)$	0
8	'Few urns'	3	18	{40,60,80}	$N(0,0.03)$	$N(0,1)$	0.0291

Note: $N()$ indicates a normal distribution and $Bern()$ indicates the Bernoulli distribution. f_g in Example 4 is a discrete distribution where μ_g is equal to 0 or 0.25 with equal probability.

I repeat each simulated example 1000 times, drawing a new set of X_{igu} each time. For Examples 1-4, which have small peer groups, I perform 100 iterations of the split sample test per simulated sample. For Examples 5-8, which have larger peer groups, I perform 800 iterations of the split sample test. For clarity, I use the word 'repetitions' when referring to the 1000 repetitions of the simulation and 'iterations' when referring to the number of iterations of the split-sample test that are performed with each simulated sample.

In each repetition r I collect $\widehat{\beta}_{GKN,r}$ and $\widehat{\beta}_{naive,r}$ as well as their parametrically estimated variances: $\widehat{\sigma}^2_{GKN,r}$ and $\widehat{\sigma}^2_{naive,r}$. In Examples 1-6, $\widehat{\sigma}^2_{GKN,r}$ and $\widehat{\sigma}^2_{naive,r}$ are clustered at the urn level. Although there is no clustering coded into the simulations, mechanical negative correlation creates its own clustering. In Examples 7-8 there are only four urns so I use heteroskedastic-robust HC2 standard errors which performed somewhat better than the clustered standard errors.

In each repetition r I collect the vector of B estimates $\widehat{\beta}_{split,b,r}$ and calculate $\widehat{\beta}_{split\_mean,r}$ as its mean. I also collect the vector of B parametric estimates of the variance of $\widehat{\beta}_{split,b,r}$, $\widehat{\sigma}^2_{split,b,r}$. $\widehat{\sigma}^2_{split,b,r}$ is estimated using heteroskedastic-robust HC2 standard errors. $\widehat{\sigma}^2_{iter,r}$ is estimated non-parametrically as the variance across the B estimates $\widehat{\beta}_{split,b,r}$. $\widehat{\sigma}^2_{split,b,r}$ will, in general, be larger than $\widehat{\sigma}^2_{iter,r}$ and $\widehat{\sigma}^2_{split\_mean,r}$ is estimated as the difference between the mean of the former and the latter.

Finally, in each repetition r I collect $\widehat{F}_r$, the F-statistic from the test of joint significance.

4 Results

Table 2 summarizes the results of this simulation.[5] The first column refers to the GKN test, the second column refers to the split-sample test, the third column refers to the permutation test, and the fourth column refers to the F-test of joint significance. A checkmark in the rows marked 'Correct size' indicates that a test at the 5% level rejected a true null hypothesis in 4-6% of the simulations. The rows marked 'Power' shows the percent of simulated samples in which the false null hypothesis is rejected at the 5% level. The rows marked 'Mean $\widehat{\beta}$' shows the average $\widehat{\beta}$ after 1000 repetitions; in the split-sample test this is an estimate of $\frac{\sigma_g^2}{\sigma_g^2+\sigma_v^2}$.

The simulations show that $\widehat{\beta}_{split\_mean}$ is a close estimate of $\frac{\sigma_g^2}{\sigma_g^2+\sigma_v^2}$. The mean $\widehat{\beta}_{split\_mean}$ of the 1000 repetitions is never more than 1% away from the true value as defined by the parameters of the data generating process. $\widehat{\sigma}^2_{split\_mean}$ is also a close estimate of the variance of $\widehat{\beta}_{split\_mean}$, as evidenced both by its similarity to the non-parametrically estimated variance of $\widehat{\beta}_{split\_mean}$ and by the correct rejection rates under the null.

The split sample test, the permutation test and the F-test of joint significance of peer group dummies all have correct size and similar power against an alternative hypothesis of positive correlation.[6] Interestingly, the F-test appears to perform well even in a situation where the assumptions of the model are deliberately not met: small peer groups with non-normally distributed characteristics (Examples 3 and 4, the 'Bernoulli pairs' examples).[7] If the researcher is primarily concerned with demonstrating that peers are not positively correlated this is a robust and easy-to-implement test; however, as mentioned previously, this test is only designed to detect positive correlations and will not reject if peers are negatively correlated.

The power of the GKN test is considerably lower than the other three tests. In the simulations where peers are positively correlated the GKN test rejects at between 1/10 and 1/2 the rate of the other tests. Furthermore the size is incorrect in several of the examples. In Example 7 I replicate a result from Wang (2009) which shows that the GKN test does not consistently eliminate mechanical negative correlation. The distribution of $\widehat{\beta}_{GKN}$ from 15,000 simulated samples shows a negative mean, statistically significant at the .0001 level.

Finally, it should be noted that as peer groups grow in size it becomes more difficult to estimate the variance of $\widehat{\beta}_{split\_mean}$ using the methodology described in Section 2. As the peer groups grow in size, $\widehat{\sigma}^2_{iter}$ converges to $\overline{\widehat{\sigma}}^2_{split}$. Since the standard error for $\widehat{\beta}_{split\_mean}$

[5]More detailed results can be found in the online appendix.

[6]The permutation test will have correct size by definition in simulations.

[7]In results not shown here, the F-test also performs well when peer groups are small and X has an exponential or uniform distribution.

Table 2: A comparison of different tests for random assignment to peers

Example 1: GKN example, random, $\sigma_g^2/(\sigma_g^2+\sigma_v^2)=0$				
	GKN	Split	Permu	F-test
Correct size	✓	✓	NA	✓
Mean $\widehat{\beta}$	-0.0008	-0.0002	-0.4841	NA
Example 2: GKN example, correllated, $\sigma_g^2/(\sigma_g^2+\sigma_v^2)=0.0291$				
	GKN	Split	Permu	F-test
Power	7.3%	60.4 %	60.5%	59.4%
Mean $\widehat{\beta}$	0.0005	0.0293	0.0101	NA
Example 3: Bernoulli pairs, random, $\sigma_g^2/(\sigma_g^2+\sigma_v^2)=0$				
	GKN	Split	Permu	F-test
Correct size	✓	✓	NA	✓
Mean $\widehat{\beta}$	-0.0005	-0.0020	-0.0254	NA
Example 4: Bernoulli pairs, correllated, $\sigma_g^2/(\sigma_g^2+\sigma_v^2)=0.0889$				
	GKN	Split	Permu	F-test
Power	11.4%	58.9%	62.5%	63.4%
Mean $\widehat{\beta}$	0.0041	0.0875	0.0641	NA
Example 5: School, random $\sigma_g^2/(\sigma_g^2+\sigma_v^2)=0$				
	GKN	Split	Permu	F-test
Correct size	X	✓	NA	✓
Mean $\widehat{\beta}$	-0.0008	-0.0002	-0.4841	NA
Example 6: School, correllated $\sigma_g^2/(\sigma_g^2+\sigma_v^2)=0.0291$				
	GKN	Split	Permu	F-test
Power	29.5%	78.9%	85.5%	85.5%
Mean $\widehat{\beta}$	0.0339	0.0293	-0.0021	NA
Example 7: Few urns, random $\sigma_g^2/(\sigma_g^2+\sigma_v^2)=0$				
	GKN	Split	Permu	F-test
Correct size	X	✓	NA	✓
Mean $\widehat{\beta}$	-0.0004	0.0002	-0.0218	NA
Example 8: Few urns, correllated $\sigma_g^2/(\sigma_g^2+\sigma_v^2)=0.0291$				
	GKN	Split	Permu	F-test
Power	49.1%	97.1%	98.9%	99%
Mean $\widehat{\beta}$	0.0256	0.0290	0.3186	NA

is the difference between $\overline{\widehat{\sigma}}^2_{split}$ and $\widehat{\sigma}^2_{iter}$, a low draw in the former and a high draw in the latter could result in the calculation of a negative variance. As the number of iterations performed increases both estimates become more precise and this becomes less of a concern. An alternative method would be to sample more than one individual per peer group, which will bias $\widehat{\beta}_{split\_mean}$ toward zero but make the variance easier to estimate. In my simulations this downward bias was not generally not large.

5 Conclusion

In comparing four different tests of random assignment to peer groups I find three which consistently perform well against an alternative hypothesis of positive correlation between peers: the split-sample test, the F-test and the permutation test. The F-test, however, is not designed to reject when peers are negatively correlated. The split-sample test is the only one of the four which generates an estimate of the magnitude of correllation between peers. The test which controls for the mean score in the urn of potential peers is found to be low power and does not fully correct for mechanical negative correlation.

References

Carrell, Scott E., Richard L. Fullerton, and James E. West, "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 07 2009, *27* (3), 439–464.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star," *The Quarterly Journal of Economics*, 2011, *126* (4), 1593–1660.

Eisenberg, Daniel, Ezra Golberstein, and Janis L. Whitlock, "Peer effects on risky behaviors: New evidence from college roommate assignments," *Journal of Health Economics*, 2014, *33* (0), 126 – 138.

Fletcher, Jason M., Stephen L. Ross, and Yuxiu Zhang, "The Determinants and Consequences of Friendship Composition," Technical Report, NBER Working Paper July 2013.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo, "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 34–68.

Hoxby, Caroline, "Peer Effects in the Classroom: Learning from Gender and Race Variation," Working Paper 7867, National Bureau of Economic Research August 2000.

Lu, Fangwen and Michael Anderson, "Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups," *Journal of Labor Economics*, 2015, *33* (1), 91 – 122.

Sojourner, Aaron, "Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR*," *The Economic Journal*, 2013, *123* (569), 574–605.

Stevenson, Megan, "Breaking Bad: Social Influence and the Path to Criminality in Juvenile Jails," Working Paper August 2014.

Wang, Liang Choon, "Peer Effects in the Classroom: Evidence from a Natural Experiment in Malaysia," Technical Report, Working Paper January 2009.

Concluding Comments

This dissertation includes two applied papers and one methodological paper on the economics of crime and peer effects. The applied papers both analyze ways in which the criminal justice system can increase socio-economic disparities. In a money bail system, those who are unable to afford bail are more likely to plead guilty and to plead to unfavorable terms. In juvenile incarceration facilities, those who are housed with peers with high levels of aggression or anti-societal attitudes will develop such traits themselves, thus increasing the likelihood that they will commit more crime after release. The methodological paper is designed to improve rigor in the study of peer influence.