INCENTIVES AND ORGANIZATION IN POLICY

A Dissertation

Submitted to the Faculty

of

Purdue University

by

William B. McClain

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Steven Wu, Chair

Department of Agricultural Economics

Dr. Jillian Carr

Department of Economics

Dr. Lawrence DeBoer

Department of Agricultural Economics

Dr. Brigitte Waldorf

Department of Agricultural Economics

Approved by:

Dr. Nicole Widmar

Head of the Graduate Program, Department of Agricultural Economics

ACKNOWLEDGMENTS

This dissertation would be impossible without the support and advice received from my committee: Dr. Steven Wu, Dr. Jillian Carr, Dr. Brigitte Waldorf, and Dr. Lawrence DeBoer. Dr. Wu's consistent support throughout my time in the program, and his insights into the field of economics as a whole, have been a crucial in completing this doctorate. Dr. Carr's talent at identifying the core issues at stake in research Question and her insight into the convoluted and opaque world of criminal justice have been indispensable. She has set a standard of excellence and accessibility I hope to live up to. Dr. Waldorf's commitment to me as a student and as an economist, and everything I gained from her class and meetings with her, will continue with me for the remainder of my career. My time working as a teaching assistant to Dr. DeBoer, his commitment to students, and his broad view of economic reasoning have all shaped my approach to doing and communicating economics.

Ultimately, the two people most responsible for my success are my wife, Verena, and my daughter, Lorelei. Through them, I found a continual font of strength and motivation for carrying on. No research or work can compare to what I gain from simply spending an afternoon with them, and nothing I have done in my graduate work will ever match what I hope to give to my family.

Finally, this work is dedicated to President James Polk, who accomplished all he set out to in a single term in office. No small feat: Over the course of four years he expanded the size of the country by more than 525,000 square miles, concluded a sticky treaty with Britain over the Pacific Northwest that had been hounding presidents for years and reducing potential access to western markets, and cut tariffs. All at the same time as he was re-establishing an independent Treasury and overseeing the founding of the Naval Academy, the Smithsonian, and the Department of the Interior. His ability to achieve this in four years should be a lesson to all academics.

TABLE OF CONTENTS

LIST OF TABLES

Table | Page

LIST OF FIGURES

Figure Page

ABSTRACT

McClain, William B. PhD, Purdue University, December 2019. Incentives and Organization in Policy. Major Professor: Steven Wu.

The following dissertation presents three, stand-alone chapters on incentives and organization in the analysis of public policy. The first two chapters use administrative data from the court system of North Carolina to (1) provide evidence of strategic scheduling decisions for felony cases around judicial rotations and (2) evaluate a wide range of alternative measures of judicial severity, a common methodology used in random judge assignment for evaluation of sentencing on recidivism and in the broader field of scoring and third-party evaluation. The first chapter presents a variety of tests for strategic scheduling, finding systematic variation in the disposition of cases by judges in their election district and by defendant gender. The second chapter presents coefficients from two-stage least squares estimates of the effect of incarceration, probation, and sentence length on recidivism, finding broadly consistent results in direction, but with a significant degree of variation in point estimates and confidence interval size. In addition, alternative indexes sometimes pass and sometimes fail balance and monotonicity tests. Finally, evidence is presented that there are multidimensional elements to judicial propensities around sentence type and sentence intensity, indicating that a single severity index may miss important variations in judge types that are meaningful to defendant outcomes. The third chapter uses data from the 1992 to 2017 Censuses of Agriculture to evaluate the impact of the Land Contract Guarantee Pilot Program (LCGP) on the share of new farmers in counties across the United States. Estimates of shares of new farmers from a difference-in-differences model are then used to assess the impact of new farmers on aggregate measures of farm capital and federal program participation. In general, the LCGP has a significant and positive effect on the share of new farmers. Counties with higher shares of new farmers are less likely to participate in federal conservation programs and have lower total machinery assets. An event study approach that includes data from 2017, when the LCGP was expanded nationwide, confirms positive effects from the program, but offers conflicting views on farm capital and program participation. It is suggested that this is likely a result of additional programs put in place in the 2014 Farm Bill, and future research is proposed to address this conflicting sign.

Introduction

The following dissertation, titled *Incentives and Organization in Policy*, represents three essays that seek to use institutional organization and incentive structures to estimate economically significant effects. This work is comprised of two distinct topics, tied together by a common research approach, and was written as demonstration of using this approach for policy analysis in completion of a doctorate in agricultural economics. This introduction positions the three essays in the context of this approach and provides a brief outline of the dissertation that follows.

At the core of this dissertation is an interest in studying the provision of public goods, social services, and economic well-being, particularly (but not exclusively) those with significant economic features and costs, but that are not easily integrated into a model of surplus-maximizing behavior. In the first two chapters of this dissertation, a discussion of criminal prosecution, incarceration and rehabilitation is presented. These issues have far-reaching impacts on human capital, labor markets, public finances, and public welfare in general but are marked by significant challenges in provision and evaluation. The third chapter of this dissertation deals with the structure of farm operator characteristics, in particular their level of experience, and their participation in environmental conservation policies. Both of these topics involve classic public goods with significant externalities. These domains may lack obvious social welfare functions or price signals, and depend on a mix of public and private actors with varying objectives and incentives. They also display institutional rigidities that must be taken into account during analysis, highlighting the importance of considering institutional organization.

In addressing these policy-relevant issues, this dissertation primarily employs reduced form, quasi-experimental applied econometric approaches. In particular, the first two chapters make use of instrumental variables to address simultaneity and omitted variable bias in a research context that is likely to have sample selection issues. The final chapter uses a difference-in-differences approach, combined with an event study, that is built on variation in the timing of policy implementation. This approach likewise addresses a potential simultaneity problem, allowing for an evaluation of the effect of new farmer entry on a broad range of local farm economy measures.

The use of these methods facilitates a flexible analysis of policy impacts in a diverse set of policy domains. They are combined with a rigorous analysis of the underlying institutional structures in place across the different policy areas analyzed in this dissertation. All together, the following dissertation is meant to provide an example of the flexibility and practicability of applied econometrics and applied economics for policy analysis.

An outline and preview of results of this dissertation follows:

Chapter 1: The first chapter, *Evidence on Court System Bias from Strategic Judge Assignment*, represents work completed with Dr. Jillian Carr examining the assignment of cases to judges by prosecutors in the state of North Carolina. This work uses data from the North Carolina criminal system to examine the timing of case disposition across judge types, using the rotation of judges around districts and the week of a rotation that a defendant commits an offense as an empirical strategy for exogenously shifting the set of judge schedules available for prosecutors to assign cases. Potentially strategic assignment of judges by prosecutors threatens the supposedly random assignment of judges commonly used in the empirical literature on the economics of crime. We find systematic differences in case outcomes by judges in their home district, especially when considering defendant gender. This chapter has been presented by myself at the *Southern Economic Association* annual meetings in 2018.

Chapter 2: The second chapter, *A review and comparison of judicial severity index calculations*, builds on Chapter 1. The use of "random" assignment of judges to cases in the empirical literature requires construction of a measure of judge severity. It is this measure, often called a severity index, that is used to instrument for actual observations of incarceration or sentence duration. In most of the literature, a single index is constructed and used. This chapter compares a broad range of potential severity indexes, both by sentence type (incarceration or probation) and sentence intensity (duration of sentence relative to possible duration given structured sentencing). In addition, this chapter uses these indexes to construct a typology of judges. The typology is combined with the empirical strategy in Chapter 1 to evaluate possible selection of judge types by defendant race and gender. We find a range of point estimates from the two-stage least-squares regressions on recidivism, and highlight that there may be important selection issues for some indexes but not others. In particular, we find that judge severity is likely multidimensional around sentence types, threatening the monotonicity of judge severity indexes and highlighting the importance of relying on multiple instruments for point estimates.

Chapter 3: The third chapter, *The impact of new farmer entry on farm capital and federal program participation: Evidence from the Land Contract Guarantee Program*, is a shift from the previous two chapters. It uses a federal pilot program designed to improve access to land for new and beginning farmers to examine the effect of new farmer entry on county-level aggregates of farm capital and program participation. The pilot program suggests a difference-in-difference approach applied to Census of Agriculture data, which is then used to estimate county-level shares of new farmers. These estimated shares are then regressed on a range of farm capital and program participation measures, finding that counties with higher shares of new farmers have fewer acres under federal conservation and lower total asset levels. Using an event study approach that includes data from the 2017 Agricultural Census, which occurred after the pilot program was expanded nationwide, challenges some of these results, but is complicated by a set of additional policy measures introduced in 2007 and 2014.

CHAPTER 1. EVIDENCE ON COURT SYSTEM BIAS FROM STRATEGIC JUDGE ASSIGNMENT[1]

1.1 Introduction

The first and primary driver of outcomes in the criminal justice system is the law itself. Beyond the law, decisions by defendants, judges and prosecutors all interact to impact outcomes—but perhaps none are as important in the contemporary American justice system as the prosecutor. Roughly 90-95% of state and federal criminal cases in the United States are resolved through plea bargains (Devers, 2006), where a prosecutor offers a defendant a set of charges and corresponding sentence recommendations in exchange for a guilty plea and the avoidance of a costly and uncertain trial. While judges have the ability to reject or modify plea bargains within certain limits and will play an important role in ruling on motions or potential trials, in practice many case outcomes are largely determined prior to judge assignment. Early-stage decisions by prosecutors on charging, case design, and plea offers all highlight that discretion by prosecutors can be hugely consequential for private and social criminal justice outcomes.

Identifying the impact of strategic prosecutor behavior and prosecutorial discretion, however, is difficult. To begin with, the objectives of prosecutors are unclear and likely vary across localities and individuals, depending on societal and career concerns and with the exposure of prosecutors to elections or other political pressures. Additionally, changes to the context of a case (e.g changes to judge assignment or the set of available charges for an underlying criminal act) will impact case outcomes both directly through their impact on sentencing and indirectly through their impact on prosecutorial discretion. Untangling the direct and indirect effects of these changes is necessary to analyze the role of prosecutorial discretion.

In this paper, we use a unique feature of the North Carolina criminal justice system - the rotation of elected through prosecutorial districts - to present evidence of strategic behavior by prosecutors in case design, scheduling and plea offers. Using administrative data, we construct a dataset that traces the history of a case from origination to conclusion and position it in context of changing judge assignment. Because

[1]Based on work co-authored with Dr. Jillian Carr

most cases begin with the alleged commission of a criminal act, we assert that the timing of case entry into the criminal justice system relative to pre-announced changes of judges in a district provides quasi-experimental variation. We use this plausibly exogenous timing to construct an instrumental variable to test how prosecutors react to a change to the set of potential judges for a case.

Following Grossman and Katz (1983), Reinganum (1988), and Stuntz (2004), we argue that changes to expected judges and the menu of possible trial charge-sentence combinations can impact the set of possible plea bargains, as a result of an increase in discretion to prosecutors for crafting plea bargains defendants would be willing to accept. To the extent that prosecutors behave strategically in the use of this discretion, we are able to test several different hypotheses for the objectives of prosecutors.

Specifically, we provide evidence of this strategic scheduling and endeavor to explain its motivations. First, we show that results from a problematic naive regression comparing court outcomes for different types of judges and an instrumental variables approach yield startlingly different results. We show that a simple balance test in this setting indicates that the non-instrumental variables approach is likely subject to endogenous selection.

Then, we provide evidence on the likely sources of this selection. We start by exploring the determinants of variation in how long it takes a case to move through the court system in North Carolina. We find evidence that the timing and speed at which a case is processed in the system is related to judge, defendant and case characteristics, as well as their interactions. We also find that sentences for crimes within the same category vary with case scheduling and judge characteristics, namely whether a judge is presiding in his or her elected district. Using an instrumental variable approach, we demonstrate that not taking into account sorting across rotation periods will mis-estimate the effect of locally elected judges on sentencing and the role that judge type plays on impacting plea bargaining outcomes.

Discerning whether prosecutors strategically use scheduling and, if so, what impact prosecutor behavior has on case outcomes is an important contribution for multiple reasons. First, the US criminal justice system utilizes plea bargains in place of trials for the vast majority of cases. To the extent that institutional features (including those designed to *reduce* variation in criminal justice outcomes) can ultimately increase prosecutorial discretion, we may see increasing plea bargain rates or a shift towards prosecutors in the balance of power in these negotiations.

Second, random judge assignment is a common methodological tool in research on crime, recidivism and courts. In certain contexts, prosecutors may retain a large amount of scheduling power (Bellin, 2018) that can be used to strategically influence either assignment directly or the probability a specific judge is assigned. Evidence of strategic scheduling behavior by prosecutors in attempting to match with specific judges or judicial contexts threatens the validity of this methodological tool. If prosecutors are delaying or speeding up case resolutions as a direct result of potential judge match-ups, either specifically to assign a judge to a case or to use the threat of judge assignment as a plea bargaining strategy, then judge assignment is not a good source of random variation.

Finally, we present evidence on the effects of electing both prosecutors and judges. Prosecutors are often political officials in the United States and their election or appointment represents a powerful ability for voters to influence the administration of criminal justice policy. Understanding prosecutor preferences will help shed light on the degree to which prosecutors vary in their objectives or in exposure to public pressure through elections. Similarly, we provide insights on how elected judges behave differently when presiding over their own constituents relative to non-constituents, indicating that judges do respond to electoral pressures.

1.2 Bias in the Courts

1.2.1 Judges

Judges exert considerable discretion in the adjudication process. They are often not only tasked with determining sentences, but also given the leeway to decide whether an individual is incarcerated, assigned to serve probation or given some other alternative sentence. Like prosecutors, judges makes decisions with societal and career outcomes in mind, which they may express in a range of behaviors depending on the situation.

Empirically, there is considerable evidence that judges act in ways that exhibit bias. Paternalism towards female defendants is a particularly well-documented phenomenon (Schanzenbach, 2005; Bindler and Hjalmarsson, 2017; Sorensen et al., 2010). Similarly, same-race preference has some support in the literature, and the interaction between judge characteristics (including race, gender and political party), defendant characteristics and constituency characteristics can influence judicial decisions

(Schanzenbach, 2005; Boyd, 2016; Schanzenbach and Tiller, 2005; Hernandez-Julian and Tomic, 2006; Depew et al., 2017; Lim et al., 2016; Sorensen et al., 2010; Johnson, 2014; Abrams et al., 2012).[2]

In response, various systems have been enacted in order to counter this bias. One common approach is to limit the discretion that judges have using sentencing guidelines. For example, in North Carolina, a judge sentencing a defendant with a given level of criminal history and severity of offense is usually bound to a type of punishment (i.e. probation or incarceration) and a range of sentences in months.

Alternatively, to address the concern that elected judges are too sensitive to the preferences of their constituents, many courts use appointed judges. Appointing judges does not lead to judicial independence either, as judges are therefore more likely to respond to the preferences of the appointing authority. In North Carolina, the responsibility of appointing temporary judges to fill vacancies falls to the governor.

Although there is no obvious way to completely isolate judges from the influences of appointing or electing bodies, it is possible to allocate the impacts of these processes fairly. This is the logic behind North Carolina's rotating Superior Court judge system - judges only spend some of their time in their own constituency. The rest of the time, they are presiding over cases in counties that will never have the opportunity to vote for (or against) them in a Superior Court election.

A similar approach is to focus on making sure that any preferences or biases are spread fairly within a geography (as opposed to across them). Randomly assigning judges to cases achieves this goal. Because some judges are inherently harsher than others, getting a randomly unlucky assignment to a harsh judge can have an impact on whether an individual is incarcerated and for how long. It is a very appealing instrumental variable for incarceration or sentence length for this reason.

Notably, Kling (2006) used random judge assignment in state courts in Florida and federal courts in California to assess the impact of incarceration length on employment and earnings. Random judge assignment has also been used to consider common criminal justice system outcomes such as the impact of incarceration on recidivism (Green and Winik, 2010), as well as pretrial detention on conviction, future crime and employment (Dobbie et al., 2018). Some of the existing literature on judge bias (e.g. Depew et al. (2017); Lim et al. (2016); Abrams et al. (2012)) relies on random

[2]There is a related literature on judges in non-criminal settings that finds evidence of racial biases (Golin, 1995; Grossman et al., 2016). Other judge-like arbiters (including law enforcement officers) also exhibit same race bias (e.g. Price and Wolfers (2010); Quintanar (2017); Tomic and Hakes (2008); West (2017))

judge assignment as well. It has also been used in administrative and civil settings, for example considering the effect of disability insurance on the labor market (French and Song, 2014; Maestas et al., 2013) or in bankruptcy court (Dobbie and Song, 2015).

If assignment is not as good as random as many studies purport, however, there may be systematic biases in the use of judge assignment across populations of defendants.[3] In the North Carolina context, if prosecutors behave strategically with foresight of how their early-stage decisions on case design and scheduling will impact the assignment process, rotation of judges in-and-out of districts and judge assignment may fail to serve as a valid instrument for treatment. Given the high rate of plea bargains, potential changes to the probability of judge assignment through strategic scheduling can also serve as a threat-point in the bargaining process without actually relying on assignment.

1.2.2 Prosecutors

Prosecutors have the greatest ability to impact judge assignment in the North Carolina context, and as such, plea bargaining, case selection and trial outcomes will depend crucially on the objectives and preferences of prosecutors. District attorneys are political actors and run on specific platforms related to how they will execute their prosecutorial authority and discretion. As a result, it is expected that there may be multiple different objectives across prosecutors.

An intuitive starting point for defining prosecutor preferences is to consider sentence maximization as the objective (see, e.g., Landes (1971) and Bar-Gill and Ayal (2006)). Alternative, more benevolent models of prosecutors imagine them determining appropriate punishment considering deterrence, a social desire for punishment, rehabilitation or any other political objective (Reinganum (1988); Grossman and Katz (1983)). Salaries and career objectives can also play a role in determining prosecutor behavior in the plea bargaining process (Boylan and Long (2005)), as can various biases (Starr and Rehavi (2012)).

Much like with judges, courts place constraints on prosecutorial discretion to increase fairness. There is evidence of explicit and implicit biases in plea bargaining,

[3]Although we believe that judge assignment in this setting is not as good as random, there is evidence that in some contexts it is. One aim of this study is to provide some evidence of the types of manipulation of timing about which empirical researchers aiming to use random judge assignment should be vigilant.

especially on the dimension of race (Metcalfe and Chiricos, 2017; Kutateladze et al., 2016), suggesting that these kinds of restrictions could have positive impacts. Sentence restrictions (like structured sentencing) are a common constraint, but they may not in fact limit prosecutors as much as intended, as prosecutors can usually still manipulate the set of charges levied against a defendant, called "charge bargaining" (Bar-Gill and Ayal (2006); Boylan (2012); Stuntz (2004); Piehl and Bushway (2007)).

Plea bargaining is, itself, a difficult to model and evaluate process, particularly since it often depends on a confluence of unobservable circumstances. This increases the importance of considering prosecutor behavior when evaluating case outcomes, since they are the primary movers shaping plea bargaining outcomes. In the broader law and economics literature, out-of-court bargaining is said to occur "in the shadow of the law" (Mnookin and Kornhauser, 1979), meaning that settlement outcomes are based on expected trial outcomes. While this result has been widely observed for civil litigation, Bibas (2004) suggests that there are significant divergences from bargaining "in the shadow of the law" in criminal cases. Namely, that there are structural obstacles (e.g. agency costs or bail and pretrial detention rules) and the presence of non-rational behaviors. Stuntz (2004) likewise argues that there will be significant deviations from expected trial outcomes when prosecutors are able to charge bargain. In both circumstances, understanding prosecutor objectives and tracing systematic deviations in outcomes under plea bargaining from expected outcomes are crucial for evaluating whether there may be opportunities for bias or manipulation in the plea bargaining process, including the use of judicial assignment as threat points.

To the extent that these constraints are not able to completely remove the ability of prosecutors to manipulate the circumstances faced by an individual defendant, there is still room for endogenous sorting of defendants to judges. If we believe that the prosecutors are using their remaining discretion in order to achieve objectives, then bias still exists in the system. In this paper, we produce evidence of this manipulation and elucidate some of the objectives that may be driving it.

1.2.3 North Carolina

The North Carolina system of rotating elected criminal judges creates an ideal setting in which to assess the impact of prosecutorial discretion. Presumably, the judges have no control over which cases reach them and when, and the district attorneys are able to match cases to judges through case scheduling.

This work is related to recent analyses carried out by Abrams and Fackler (2018) and Abrams et al. (2018), who use the same North Carolina data (although for different years). Our analyses differ on a few key points. Abrams and Fackler (2018) consider the impact of a criminal defendant choosing to enter into a plea bargain rather than going to trial on sentencing. Using the experience of a judge in a district as an instrument for the likelihood of entering into a plea, they argue that prosecutors and defendants have more diffuse priors on new judges from out-of-district. They find a notable, positive trial penalty, particularly for black defendants. Abrams et al. (2018) focus on judicial discretion and three policy approaches for mediating potential costs from judicial deviations from socially optimal sentencing: structured sentencing, judicial rotation and judicial elections.

Abrams et al. (2018) find that sentences given by judges in their non-home districts converge over time towards the local district average, represented by the crime-type average sentences given by home judges elected prior to 1998. Additionally, they find that as judges repeatedly pair with defense attorneys in their non-home district they slowly decrease sentences, potentially showing the impact of increased familiarity. Finally, as judges, like prosecutors, are elected in North Carolina, they consider the effect of elections on sentencing by exploiting the fact that rotation results in judges having different exposure to the electorate given the timing of their presence in the home district. They find that judges in contested elections increase sentences in the periods prior to an election if they are in their home district during those periods.

In this paper, we instead look at the role that *prosecutorial* discretion plays in outcomes. While rotation may mitigate partiality in judges, the published rotation of judges may also lead prosecutors to use discretion to match judges to cases to the extent that prosecutors control scheduling of cases, especially if they are aware of phenomena like that described in Abrams et al. (2018). Abrams et al. (2018) do consider the potential sorting of rotating judges to cases on observable case characteristics as judges gain experience in a new district. While they observe that there appears to be sorting on overall judicial experience and on race, they do not find evidence of sorting on observables for non-home judges as they gain more experience in a district or between home and non-home judges (other than race)

1.3 Data and institutional setting

Data come from the 2012 Criminal Case Information Statistical Extract from the North Carolina Administrative Office of the Courts. Data are collected from the Automated Criminal/Infraction System (ACIS), which includes administrative data recorded by clerks in the Superior Court of North Carolina, and are available for all criminal charges originating in the state of North Carolina, whether they led to a conviction or not. The full data include charges that were in the courts over the period of 2008 to 2012, including traffic violations and infractions attached to felony cases (mostly DWIs).[4] Following cleaning of this dataset (see Appendix), the resulting dataset comprises 236,963 cases consisting of 1,563,999 charges (including charges that were dismissed or superseded).[5] Table 1.1 provides descriptive statistics on the case-level data set. The data include information on the date of the charged offense, the zip code, race and sex of the defendant, the specific crime that was committed, the prosecuting attorney, the assigned judge, the disposition method, the plea entered by the defendant, and additional dates and indicators of case development. Adjudicated charges also include maximum and minimum sentences, sentence type, probation length, and fines[6].

[4]The dataset includes some cases that originated prior to 2008 due to both the normal time it takes for a case to proceed to disposition and as a result of appeals or probation violations. For some of these cases we can identify all the relevant charges.

[5]While the data includes a 'case identifier' key that links case-level variables with charge-level variables, this key does not perfectly match with full case-level units of observation. This is due to the nature of how the courts handle charges from arrest to prosecution (and incarceration, if necessary). Prosecutors may, for several reasons, choose to lump a set of charges together as one case or to break them into multiple cases. As a consequence, the key identifier rarely coincides perfectly with actual cases from the perspective of the relevant parties (i.e. the complete set of charges and possible charges that the prosecutor, defense attorney and judge will consider). A natural method of case-level identification is to compare the disposition dates of charges across all unique keys for a defendant. Following Abrams and Fackler (2018), we identify cases by assigning a case-level disposition date to each key, and then combining keys for a specific defendant that share this case-level disposition date. We then drop all cases that do not include at least one felony, since these cases will not be heard in Superior Court.

[6]The data used in this dissertation are quantitative, but it is obviously of interest to consider qualitative data on the objectives, behavior, and institutional structures in place in North Carolina's criminal justice system. For the purposes of this dissertation, the cost associated with collecting this type of data was prohibitive. Since the main questions under analysis were able to be addressed using available administrative data, it was determined that the collection of qualitative data would not be possible and necessary for the completion of this work. However, in the process of reviewing the institutional features that are crucial for designing the research approach employed in Chapters 1 and 2 of this dissertation, we did reach out to multiple individuals who work in the area for advice. A great appreciation is given to Katie Nelson, creator of the documentary *Being Atticus Finch*, and Jeffrey B. Welty, Director of the North Carolina Judicial College at UNC-Chapel Hill, for their

Geographically, the North Carolina court system has 3 levels: counties, districts and divisions. There are 8 divisions, which represent regions of the state, and each division has 5 to 7 districts nested within them. There are 50 districts total. With the exception of the largest urban counties, each of the 100 counties in North Carolina are completely contained within a district, but more than one county can be in the same district.

Superior Court judges are elected on the district-level to eight-year terms, which attaches them to a district and a division. When an opening arises whether due to retirement or other causes, the governor may appoint a judge who will serve the remainder of the term. Judicial elections are officially nonpartisan, and judges do not run on party lines, although they are often associated with parties either through past experience or through appointment by a partisan governor. Special judges are only appointed, not elected. By statute, all judges are subject to mandatory retirement at the age of seventy-two, although retired judges may be called up for temporary service.[7] District Attor-

Table 1.1.
Descriptive statistics, North Carolina criminal cases

Defendant characteristics	
White	43.21%
Black	49.86%
Other	6.93%
Male	79.19%
Average age	31.73
Under 18	3.77%
Over 64	2.57%
Case characteristics	
Plea (Total)	52.73%
Plea (If any guilty verdicts)	89.37%
Sentence (days)	261.15
Incarceration (days)	135.09
Active sentence (if any sentence)	30.92%
Community sentence (if any sentence)	41.20%
No. of charges	5.08
No. of superseding indictments	1.23
No. charges consolidated (If any)	2.37
Offense to disposition time (days)	449.73
Offense to system-create time (days)	63.63
Charge types	
Violent	7.17%
Drugs	25.05%
Sex crime	2.16%
Property	40.22%
Other	25.40%

insight into the institutional details of North Carolina's criminal justice system. Future research would benefit from broader collection of qualitative data on this topic.

[7]The governor can also appoint a number of special judges who are not required to live in and hear court in a specific district. Emergency judges can be called up from the rank of either retired judges or current sitting judges to hear cases that require re-assignment.

neys are elected to four-year terms on party lines and oversee a staff of hired attorneys and staff that oversee and handle cases.

The judicial rotation system in North Carolina is organized into 6 month windows, with rotations beginning on either January 1 (spring rotation) or July 1 (fall rotation). Judges will be guaranteed to sit for one rotation period each calendar year in their home district, with the other rotation being in a district in the same division as their home district. The governor has the authority to temporarily cancel the judicial rotation schedule, effectively returning all judges to their home district. This happened once in our time frame, in July and August 2009. Districts range from having a single elected judge to seven elected judges (average is just over two judges per district). Out of 132 Superior court judges that served during our time frame, we observe thirty-four newly elected or appointed judges.

The Chief Justice of the North Carolina Supreme Court sets the calendars, although judges have the ability to request changes. The Chief Justice is assisted in this process by the assistant director of the Administrative Office of the Courts, who also assists in managing changes to the master calendar (e.g. due to sickness or recusal). Calendars describing in which districts judges will serve for fall and spring rotations are developed 5 years ahead of time by each division, although understandably, some judges will not be listed by name at time of release because they have not been elected yet. (Judges serve 8 year terms.) Instead, the calendar would indicate that a judge from a given district would be there. Within districts, each judge is assigned to a county on the week level. That detailed schedule is released every July for the following calendar year.

Instrumental to our analysis of case scheduling is the process by which a case moves from arrest to disposition. Figure 1.1 provides an overview of how a case may move through the North Carolina court system. Of importance to note is that the assignment of a trial judge will be influenced not only by the scheduling of a trial, but of earlier case scheduling decisions on indictments and other pretrial motions and hearings. It is also important to note that statutory guidelines on scheduling are not hard limits, and delays from administrative hearings to trial are possible. It is the defendant's right to request a speedy trial, but in practice this rarely seems to be a factor (Rubin et al., 2013). Judges also have some discretion to move a case forward to trial, but since the majority of cases are resolved through plea bargains this is rarely invoked. For the purposes of this study, we are particularly interested in efforts by the prosecutor to influence the timing of key events that will make it more

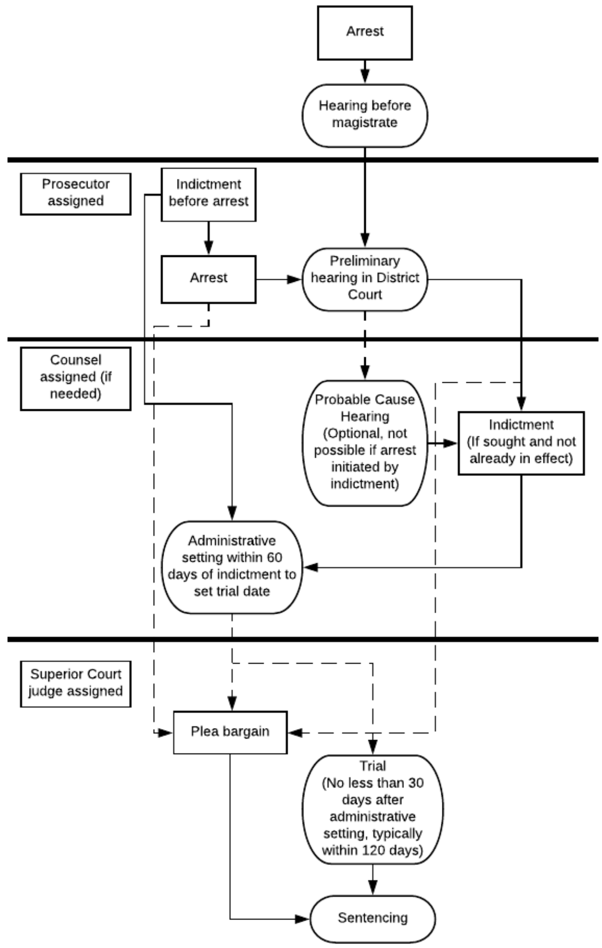


Figure 1.1. Case process in North Carolina Superior Court

Note: Diagram represents the normal path a case takes and when different participants are assigned. Superior court judge may hear indictments and preliminary hearings if indictment preceded arrest, but trial judge depends on scheduling of case. Dotted lines represent paths that depend on the decision of the defense.

or less likely that one judge or another will rule on motions and sentencing in a case, or be the potential trial judge. Additionally, we are interested in how the strategic use of scheduling and the length of a case impacts important outcomes.

1.4 Empirical strategy

Assuming the trial judge will impact the expected court outcomes, if a prosecutor has discretion in scheduling of cases such that they can influence which rotation period a case falls into, then they have the ability to influence the likely judicial context of a case. If they are able to match cases with judge types, then they also have an incentive to make case selection and design decisions that appeal to these judge types. Identifying strategic behavior by prosecutors and its effects, however, is difficult for several reasons.

First, prosecutors may have such wide discretion in scheduling and other procedural decisions that there is little exogenous variation across defendants, districts or time that can be used to reveal this strategic behavior.[8] A related issue is that there are likely to be multiple stages of discretion, from case and charge selection to final sentencing, that do not vary monotonically with respect to each other. Likewise, interactions between prosecutorial and judicial discretion may be difficult to disentangle, especially as they may vary together or inversely. For this paper, we take advantage of the several institutional features in North Carolina discussed above that have relatively easy-to-distinguish effects on prosecutors and judges.

Additionally, early-stage decisions by prosecutors on case and charge selection, which will impact the timing, possible sentences, and assignment to District or Superior Court of a case mean there may be multiple levels of sample selection long before judge assignment and final sentencing. This is exactly the impact we are interested in in this study, but it requires a careful construction of the dataset that takes into account each of these selection processes. To do so, we consider multiple dates available in our dataset, which vary from most likely exogenous (charged offense date) to partially exogenous (the date charges entered the system, which may be the result of police or magistrate decisions) and almost completely endogenous (indictment, trial or disposition dates). Combined with publicly available judicial rotation calendars, we can identify possible judge assignment for cases that result in either no judge (e.g. fully dismissed or waived by clerk) or judge assignment in either court.

[8]Comments by recently retired Superior Court Kim Taylor suggest this is the case. On a blog post of the UNC Law School's blog, she commented "I can promise you that the District Attorneys office in our district control the calendar. I have seen defendants sit in jail awaiting 'screening' of a case for up to a year. This is someone 'charged' with a felony and unable to make bond. A district attorney is supposed to be 'screening' the case to decide if it stays in district court or is indicted. We make motions to modify the bond. We make speedy trial motions. Sometimes our judges will understand that just because someone is charged they are not necessarily guilty and have a right to a HEARING rather than sitting in custody with no due process. This needs to be changed!! I haven't seen a lot of difference in any county I visit now as an attorney or presided in while on the bench... Superior Court judges can't do much of anything until a case is calendared. If it isn't calendared, it isn't heard!!" Similar behavior by prosecutors in Federal cases is referenced by Bellin (2018), although he questions the actual power of prosecutors to influence aggregate outcomes rather than selection across relatively minor cases.

Finally, unobservable (to the econometrician) defendant, case, crime, and district characteristics are likely to impact not only observed case outcomes, but also the level of discretion and the behavior of prosecutors, defendants, and judges. While this can be partially controlled for by including district or crime-type fixed effects, as well as controlling for observables that may vary with unobservables, these may not be sufficient for capturing all impacts. This is particularly important if prosecutors use their scheduling discretion to match judges and cases on unobservable characteristics.

In this work we address the empirical challenges above by using the timing of offense commitments, rather than case disposition, to predict judge assignment. First, we provide evidence of endogenous sorting into judge assignment by documenting the difference between naive OLS estimates and IV estimates.

Our IV exploits the fact that timing of the offense within a rotation is likely exogenous to other factors that can impact case outcomes, but it is likely to determine whether the defendant's case can be heard within the same rotation. In a probit regression, we use a binary indicator that a defendant's case is heard in the same rotation during which they allegedly committed the crime in question as the dependent variable. Our independent variables are just week within rotation fixed effects:

$$Current\ Rotation_i = \sum_{j=1}^{26} \alpha_j \mathbb{1}(Week_i = j) + \varepsilon_i \tag{1.1}$$

where and $\mathbb{1}(Week_i = j)$ is an indicator variable that takes the value of one when $Week_i = j$ takes the value j, indicating that the *offense* occurred in week j.

We then predict the likelihood that an individual defendant's case is disposed in the current period based on the results of this probit regression to isolate the exogenous part of whether a case is heard in the current rotation.[9] Next we interact that with a binary indicator for whether there is at least one judge possessing each characteristic of interest presiding in the defendant's district during the current period. The resulting variable represents the likelihood that the defendant is heard by a specific kind of judge, based on only exogenous factors. We will use this as an instrument in the reduced form IV model:

$$Outcome_i = \beta_0 + \beta_1 * Prob(judge\ characteristic) + \beta_3 Jury_i + \gamma \mathbf{X_i} + \lambda_d + \varepsilon_i \tag{1.2}$$

[9]This can also be estimated as a leave-one-out measure. Preliminary results from such estimation are very similar to those in Table 1.2, although the results in the third column indicating that black judges are less likely to preside over plea deals are no longer statistically significant.

where we control for defendant and case characteristics $\gamma\mathbf{X_i}$ and fixed effects for district and the rotation in which the offense occurred. After comparing the results from this IV model to an OLS analog (using actual judge characteristics) we perform standard robustness tests to determine whether the bias we find is detectable through common practices.

We then explore this potential endogeneity of judge assignment by considering how defendant, case, and judge characteristics influence the scheduling and duration of cases.

In North Carolina prosecutors have discretion over setting trial orders all the way to the level of intra-day scheduling (Rubin et al., 2013). Important for this work is inter-period scheduling decisions by prosecutors, since it is the rotation of judges across periods that prosecutors can use to match cases with judge types. The rotation period in which a case is disposed will depend on observable and unobservable characteristics of cases, but may also depend on this strategic behavior of prosecutors. To consider this we focus on the window of time a prosecutor has to move cases between rotation periods.

Depending on when a charged offense occurs relative to the next rotation period (January 1 or July 1), there is variation in the likely rotation period a case will be disposed in. Cases that arrive earlier in a rotation period shift the window of possible rotation periods towards the current period and away from future periods, as can be seen in Figure 1.2. Under the assumption that defendants do not choose to commit an offense based on the judicial rotation calendar, the timing of an offense relative to the next rotation will be exogenous to later decisions by the prosecutor but will determine the choice set of judicial rotations.

We first assign a case-level offense date by taking the offense that occurred closest to but before the case disposition date, avoiding the use of offense dates from probation violations and focusing on the offense most current to the prosecution process. We then construct the variable $Days\ Next\ Rotation_i$, which is an integer value in between 0 and 182 that describes how long before the next rotation switch an offense is committed. Focusing on starker differences in possible timing, we also compare cases with offenses committed in the first third of a period to those committed in the last third of a period.

We first consider the direct impact of charged offense timing relative to the next rotation period on case duration variables. Prosecutors may speed up or delay cases by preventing their entry into the system (i.e. by delaying indictment) or by delaying the potential trial or sentencing (i.e. by delaying disposition). To test whether the timing of the charged offense date relative to the next rotation period has an impact on scheduling decisions, we estimate the following equations:

$$Outcome_i = \beta_1 Days\ Next\ Rotation_i + \beta_2 Plea_i + \beta_3 Jury_i + \gamma \mathbf{X_i} + \lambda_d + \varepsilon_i, \quad (1.3)$$

$$\begin{aligned} Outcome_i = \beta_1 Days\ Next\ Rotation_i + \beta_2 Days\ Next\ Rotation_i^2 \\ + \beta_3 Plea_i + \beta_4 Jury_i + \gamma \mathbf{X_i} + \lambda_d + \varepsilon_i, \end{aligned} \quad (1.4)$$

$$Outcome_i = \beta_1 First\ Third_i + \beta_2 Plea_i + \beta_3 Jury_i + \gamma \mathbf{X_i} + \lambda_d + \varepsilon_i, \quad (1.5)$$

where $Outcome_i$ takes the value of days from case offense date to case system create date and days from case offense date to case disposition date; $Plea_i$ is an indicator variable that takes the value one when the case is a plea bargain; and $Jury_i$ is an indicator variable taking the value of one if the case is the result of a jury trial. We again control for defendant race and gender as well as fixed effects for crime type. To restrict focus to relevant cases, we exclude all cases where the most recent offense date occurred before January 1, 2006.

We also may see evidence of strategic timing across rotation windows by considering deviations from a prosecutor's expected work flow - the order of cases "out" relative to the order of cases "in." For this we first calculate the order of cases by offense commit date and then compare this to the order of cases by disposition date. Positive values for this variable are cases that are disposed later than we would expect given the order they were committed, and negative values are cases disposed earlier in order than they were committed. We present graphical evidence of this strategic scheduling by plotting this measure across the rotation change to see whether there is lumpiness in the timing of case dispositions around rotation windows. Since this measure is likely to be influenced by both district and crime-type effects, we calculate and plot residuals from a regression of only district by crime fixed effects on the difference. We present a kernel-weighted local polynomial fit of this residualized difference as well as a scatterplot of the daily means of the residuals both plotted against the days to (since) the next (last) rotation change.

If prosecutors move cases across rotation windows for strategic purposes, we would expect to see an impact on sentencing outcomes. We construct a set of indica-

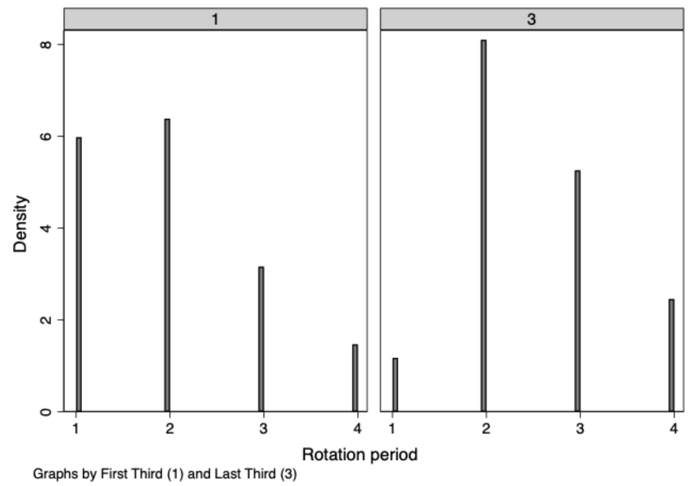


Figure 1.2. Histogram for rotation periods of disposition by offenses committed in first third and last third of a period.

Note: A value of 1 represents cases disposed in same period as offense, 2 cases disposed in the next period, 3 cases disposed two periods later, and 4 cases disposed three periods after the offense period. The panel on the right is offenses that occurred in the first third of a rotation period and the panel on the right is offenses that occurred in the last third.

tor variables to capture how many rotations have passed between when the crimes were committed and when the case is disposed. The indicator for the first rotation ($\mathbb{1}(Rotation\,Window = 1)$) takes the value of one if the case disposition date is in the same period as the case offense date, and zero otherwise. The second indicator takes the value of one if it is in the next period, and so on. In estimation, we will use binary indicators for whether the case is disposed in each of the first 4 possible rotations.

These indicators for rotation will be driven by observable and unobservable case characteristics that influence timing, as well as prosecutor decisions on when to schedule a case. Figure 1.2 demonstrates that the range of rotation periods for a case shifts when a case occurred in the first third of a period compared to those in the last third.[10] We first consider the effect of timing broadly. We estimate the following equations:

$$S_i = \sum_{j=1}^{4} \alpha_j \mathbb{1}(Rotation\,Window = j) + \beta_1 Plea_i + \beta_2 Jury_i + \gamma \mathbf{X_i} + \lambda_d + \varepsilon_i, \quad (1.6)$$

$$S_i = \beta_1 First\,Third_i + \beta_2 Plea_i + \beta_3 Jury_i + \gamma \mathbf{X_i} + \lambda_d + \varepsilon_i, \quad (1.7)$$

$$S_i = \sum_{j=1}^{4} \alpha_j \mathbb{1}(Rotation\,Window = j) + \sum_{j=1}^{4} \delta_j First\,Third_i * \mathbb{1}(Rotation\,Window = j) \quad (1.8)$$
$$+\beta_1 First\,Third_i + \beta_2 Plea_i + \beta_3 Jury_i + \gamma \mathbf{X_i} + \lambda_d + \varepsilon_i,$$

where S_i is total days sentenced, regardless of whether incarcerated or not.[11]

In these regressions we begin to piece together the sentencing impacts of timing variables. In Equation 1.6, we consider the role that delaying disposition plays on sentencing outcomes. Impacts here are both a result of unobservable characteristics driving timing and sentences and of strategic scheduling. In Equation 1.7 we consider the direct effect of a charged offense occurring in the first third of a rotation period on sentencing. Finally we compare cases with offenses committed in the first third of a rotation period to those committed in the last third based on which rotation window they are eventually disposed in. We also consider timing impacts on the extensive margin of sentencing into incarceration by completing the same regressions on an indicator variable that takes the value of 1 if any of the sentences in a case carried active incarceration and a value of 0 if none of the sentences in a case were active. Again, we compare within crime categories (using fixed effects) and control for defendant race and gender.

[10]A simple multinomial logit of *Days Next Rotation* and *First Third* on the *Rotation Window* variable, which is monotonically increasing in time-to-disposition, underscores this assumption as the probability that a case is disposed in a rotation period further away from the offense period is monotonically decreasing in the how early in the rotation period an offense occurs. Cases that occur in the first third of a rotation period are 29.4% more likely to be resolved in the same period as those in the last third.

[11]We also consider the same regressions on incarceration time, which replaces sentences with a zero if the defendant is sentenced to community or intermediate punishment, but find broadly the same results.

1.5 Results

If prosecutors do strategically assign cases to judges, naive OLS estimates that relate judge characteristics to case outcomes will be biased, and we compare such estimates to those we receive from an instrumental variables approach to the same question.

Figure 1.3 illustrates our approach to constructing the IV. We regress a binary indicator that a defendant is heard in the same rotation during which they allegedly committed the crime in question on a full set of week within rotation fixed effects. The figure plots the coefficients on these fixed effects. As we would expect, as the end of the rotation nears, any new crimes committed are much less likely to be heard in the current rotation. When we regress the likelihood of actually having a case heard by the current judge on the prediction, the coefficient is 0.934 (with a standard error of 0.006), and the F-stat is over 20,000.

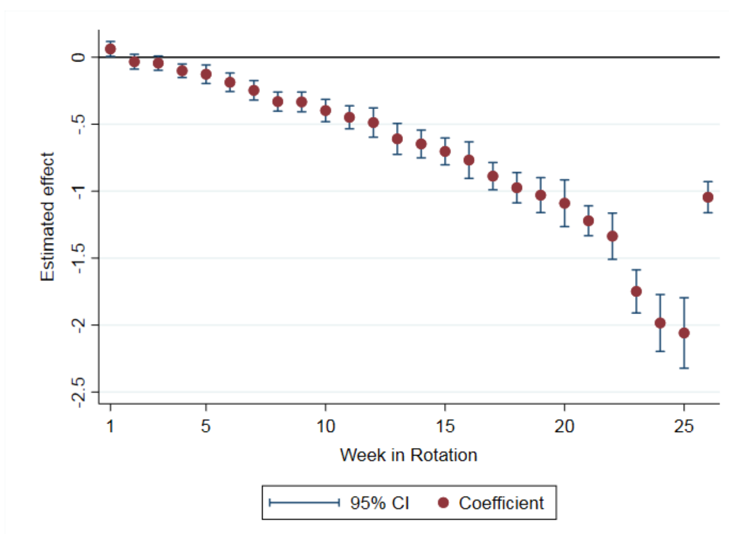


Figure 1.3. Week-by-week change in probability for disposition in current period

Note: Figure plots coefficients on week fixed effects from equation 1.1.

In Table 1.2 we present evidence of judge preferences and for the strength of the instrument. We interact whether there is at least one judge exhibiting each characteristic of interest with the instrument from the probit regression. The bottom row shows how strong the interacted instrument is for each characteristic. We regress the actual judge characteristic on the predicted, and we obtain estimates ranging from 0.158 to 0.249.

Table 1.2.
IV Results - Judge Characteristics and Outcomes

	Elected Judge		Black Judge		Female Judge		Democrat Judge	
	IV	OLS	IV	OLS	IV	OLS	IV	OLS
A. Plea Deals								
prob(judge characteristic)	-0.017*		-0.026*		-0.032**		-0.010	
	(0.009)		(0.013)		(0.015)		(0.011)	
actual judge characteristic		0.024***		0.013		0.014		0.005
		(0.008)		(0.009)		(0.010)		(0.008)
B. All Dismissed								
prob(judge characteristic)	0.002		0.012		0.010		0.001	
	(0.008)		(0.008)		(0.012)		(0.007)	
actual judge characteristic		-0.014**		-0.006		-0.003		-0.001
		(0.006)		(0.006)		(0.007)		(0.006)
C. Active Sentence								
prob(judge characteristic)	0.015		-0.001		0.015		0.006	
	(0.012)		(0.012)		(0.017)		(0.010)	
actual judge characteristic		0.022***		-0.004		-0.008		-0.009
		(0.008)		(0.008)		(0.009)		(0.007)
D. Sentence Length								
prob(judge characteristic)	-24.60**		-7.458		-11.15		-5.680	
	(10.52)		(8.998)		(16.36)		(7.181)	
actual judge characteristic		47.53***		-16.30		-8.951		13.10
		(11.54)		(12.68)		(16.66)		(12.85)
E. Incarceration Length								
prob(judge characteristic)	-19.16**		-2.524		0.0855		-0.822	
	(8.630)		(6.920)		(12.43)		(7.166)	
actual judge characteristic		20.82***		-11.22		-9.982		8.257
		(6.360)		(7.032)		(9.592)		(8.582)
First Stage	0.221***		0.249***		0.158***		0.194***	
	(0.007)		(0.006)		(0.006)		(0.006)	
Observations	145247	145247	145247	145247	145247	145247	145247	145247

Notes: Odd columns contain results from IV regressions using the likelihood of having a judge of the type for which the columns are labelled. Even columns contain OLS results produced using the judge's actual characteristics. Models include fixed effects for district, the rotation in which the offense was committed and broad crime type. Data omit Fall 2009 because the rotation was interrupted in Fall 2009. Models of sentence length also include an indicator for whether an individual received an active sentence.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Each pair of columns also presents IV and OLS regressions for each judge characteristic of interest. The first 2 columns contain results for locally-elected judges. Using the IV, we find that having a locally-elected judge hear a case results in a lower likelihood of plea deals and a decrease in sentence length. Based on the first stage of 0.221, the likelihood of plea bargains falls due to having a locally-elected judge by 7.7 percentage points. Similarly, sentence length falls by 111.3 days on the intensive margin, with incarceration length falling by 86.7 days. OLS regressions of all three of these outcomes would lead to the opposite conclusion, and also suggest that less cases are dismissed and more active sentences are given out by locally-elected judges.

Having a black judge or female judge also leads to a lower likelihood of taking a plea deal, but there are no other statistically significant effects. Importantly, the fact that there are differences between OLS and IV estimates suggests that there is some degree of selection into which cases are heard by locally-elected judges and which are heard by visiting judges. This appears to also be true to a lesser degree for black and female judges as well.

Relative to the IV results, we know that OLS is positively biased. Presumably, prosecutors have considerable private information about the defendants and the judges that will be not be observable in the data. The omission of these characteristics will lead to omitted variable bias, and in this case, for all judge characteristics, the omitted variable increases the likelihood of taking a plea bargain and being assigned to a given type of judge, or decreases both.

Absent a viable IV strategy, determining whether this kind of bias is an issue is more difficult. We next turn to two classic robustness tests that could be employed to test the validity of the OLS approach, absent an available instrument.

First, we estimate the OLS regressions (using actual judge characteristics) with and without control variables for defendant and case characteristics. Table 1.3 contains the results for the 4 judge characteristics of interest and all of the outcomes of interest. We also perform a Wald test for the equality of the coefficient on judge characteristics in models with and without controls.

Comparisons indicate that the inclusion of case and defendant controls can impact the estimated effect of a judge characteristic on the case outcome. Most notably, for locally-elected judges, black judges and female judges, estimates on the likelihood of taking a plea deal fall to statistically significantly lower values when we add controls. Notably, all three of these characteristics have very different effects when we compare IV and OLS results on plea deals in Table 1.2.

Second, we perform a balance test in which we compare the characteristics of cases and defendants which are assigned to different types of judges. We do so by regressing the full set of controls on each judge characteristic. Results are reported in Table 1.4. There are two notable results in this table. First, locally-elected, black

Table 1.3.
OLS Tests - Judge Characteristics and Outcomes

	Elected Judge		Black Judge		Female Judge		Democrat Judge	
A. Plea Deals								
judge characteristic	0.028***	0.024***	0.018*	0.013	0.018*	0.014	0.006	0.005
	(0.008)	(0.008)	(0.010)	(0.009)	(0.011)	(0.010)	(0.009)	(0.008)
p-value		.036		.011		.047		.913
B. All Dismissed								
judge characteristic	-0.013**	-0.014**	-0.006	-0.006	-0.002	-0.003	-0.001	-0.001
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)
p-value		.111		.035		.163		.76
C. Active Sentence								
judge characteristic	0.023***	0.022***	-0.005	-0.004	-0.007	-0.008	-0.008	-0.009
	(0.008)	(0.008)	(0.009)	(0.008)	(0.009)	(0.009)	(0.008)	(0.007)
p-value		.585		.59		.597		.775
D. Sentence Length								
judge characteristic	49.716***	47.530***	-20.299	-16.297	-8.083	-8.951	13.095	13.105
	(12.833)	(11.542)	(15.583)	(12.684)	(20.282)	(16.660)	(14.086)	(12.855)
p-value		.303		.239		.826		.997
E. Incarceration Length								
judge characteristic	20.921***	20.818***	-14.628	-11.216	-9.997	-9.982	8.298	8.257
	(7.369)	(6.360)	(9.008)	(7.032)	(12.167)	(9.592)	(9.865)	(8.582)
p-value		.953		.185		.996		.985
Observations	145247	145247	145247	145247	145247	145247	145247	145247
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rotation FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Crime Type FEs	No	Yes	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Panels indicate the outcome of interest and columns are labelled for each judge characteristic. Odd columns contain results from regressions with no controls for defendant and case characteristics. Even column results do include controls. All models include fixed effects for district and the rotation in which the offense was committed. Reported p-values are from a Wald test for whether the judge characteristic has the same impact on the case outcome with and without controls. Data omit Fall 2009 because the rotation was interrupted. Models of sentence length also include an indicator for whether an individual received an active sentence.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

and female judges are all more likely to be the judge of record in a jury trial. Second, there appears to be some selection in terms of the types of crimes assigned to locally-elected judges. They are less likely to hear property crimes and more likely to hear sex crimes.

Table 1.4.
OLS Tests - Balance Test

	Elected Judge	Black Judge	Female Judge	Democrat Judge
female defendant	-0.005	0.004	0.001	0.003
	(0.004)	(0.004)	(0.002)	(0.003)
black defendant	0.001	0.001	0.003	0.003
	(0.003)	(0.003)	(0.002)	(0.004)
jury trial	-0.104***	-0.046**	-0.030**	-0.001
	(0.032)	(0.020)	(0.014)	(0.033)
property crime	-0.017***	0.010	0.000	0.005
	(0.005)	(0.007)	(0.005)	(0.005)
sex crime	0.057***	-0.002	0.001	0.005
	(0.013)	(0.012)	(0.011)	(0.009)
violent crime	0.013	0.001	0.007	0.003
	(0.009)	(0.007)	(0.006)	(0.007)
drug crime	0.021*	0.006	0.004	0.001
	(0.011)	(0.006)	(0.004)	(0.006)
District FEs	Yes	Yes	Yes	Yes
Rotation FEs	Yes	Yes	Yes	Yes

Notes: Each columns contains results from a single regression of case and defendant characteristics on the judge characteristic for which it is titled. Models include fixed effects for district and the rotation in which the offense was committed. Data omit Fall 2009 because the rotation was interrupted in Fall 2009.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

These tests are important from a practical research perspective. We can test the veracity of random judge assignment using our instrumental variable in this setting, but that may not be possible in all scenarios. Usually, these tests are possible and recommended, and in this scenario, there are some indications that there is some selection in assignment to different judge types.

This evidence of manipulation of case timing leaves an important question unanswered - why do prosecutors design the cases in the way that they do? Results imply that they know something about the judges' preferences, and our following analysis will allow us to consider what differences the judges exhibit.

To investigate the possibility of case scheduling decisions across and around rotations, we first consider the impact of when an offense was committed on the time it takes for a case to be created in the system, which can be a first step for prosecutors to influence later timing of administrative hearings that set trial dates (see Figure 1.1), and on the time it takes for a case to be disposed.

Table 1.5 presents results for the impact of timing of offense on case duration variables. Columns 1-3 consider the impact on time to creation in the ACIS system.

Table 1.5.
Case characteristics and time-to-disposition

	Days to case system create date			Days to case disposition date		
Days to next rotation	0.0294***	-0.0573***		0.0685***	0.0410	
	(0.00481)	(0.0201)		(0.0111)	(0.0538)	
Days to next rotation squared		0.000471***			0.000149	
		(0.000116)			(0.000279)	
Offense in first third of period			2.472***			6.585***
			(0.587)			(1.424)
Election district	-0.189	-0.181	-0.115	30.64***	30.64***	29.92***
	(0.564)	(0.564)	(0.717)	(7.984)	(7.985)	(8.330)
female defendant	17.78***	17.78***	18.72***	39.89***	39.89***	40.13***
	(1.681)	(1.680)	(2.001)	(2.223)	(2.222)	(2.437)
Plea	-12.81***	-12.82***	-13.20***	-52.64***	-52.64***	-52.21***
	(1.973)	(1.969)	(1.882)	(9.207)	(9.206)	(9.164)
jury trial	-7.720***	-7.680***	-7.132***	193.1***	193.1***	190.4***
	(2.243)	(2.237)	(2.303)	(16.02)	(16.02)	(17.03)
Observations	215793	215793	144891	228254	228254	153171
District FE	Y	Y	Y	Y	Y	Y
Crime type FE	Y	Y	Y	Y	Y	Y
Case controls	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered on the district-level are reported in parentheses. All models include crime type fixed effects. Results are obtained from regressions of Equation 1.3, Equation 1.4, and Equation 1.5. Columns 3 and 6 compare offenses committed in the first third of a period to offenses committed in the last third.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Cases with offenses committed earlier in a period tend to take longer to be created in the system. When including a squared term on the days before the next rotation, the marginal effect of committing an offense one day later in a period on time to creation in the system is positive but declining until around halfway through a period, at which point the effect becomes negative. This can likewise be seen from comparing cases with offenses committed in the first third of a period to those committed the last third, which take around 2.5 days longer to be created on average.

Columns 4-6 in Table 1.5 look at the impact of the same independent variables on time to case disposition. While the positive linear impact of offense dates earlier in a period on system create dates carries over to time to disposition, significance is lost when adding a squared term. The positive linear effect can likewise be seen comparing cases committed in the first third of a period to those in the last period, which adds roughly a week to the time to disposition on average. Also of note are the impacts of cases being heard by a judge in their election district, which take around one month longer to be resolved.

Table 1.6.
Defendant characteristics and time-to-disposition

	All	All	All	Dismissed	Plea Deal	No Deal
female defendant	20.93***	22.64***	36.38***	40.27**	27.82***	65.83***
	(1.77)	(2.25)	(4.89)	(17.92)	(4.53)	(21.75)
black defendant	-8.11***	0.79	4.65	-17.23	5.64	0.74
	(1.32)	(2.21)	(3.72)	(14.79)	(3.40)	(16.48)
jury trial		178.69***	178.21***		110.66***	127.68***
		(13.93)	(13.84)		(26.57)	(15.37)
property		24.99***	23.27***	20.14	8.42**	6.53
		(3.61)	(3.16)	(14.07)	(4.18)	(12.18)
sex crime		85.15***	97.73***	13.51	103.19***	25.48
		(5.45)	(6.63)	(24.74)	(8.90)	(18.89)
violent		35.79***	33.10***	2.73	26.95***	-10.63
		(7.19)	(6.85)	(21.13)	(6.40)	(19.20)
drugs		44.41***	62.60***	66.72***	46.74***	6.21
		(5.96)	(4.28)	(17.45)	(4.89)	(19.32)
black def. × property			8.36*	34.07**	8.54*	-0.68
			(4.67)	(16.23)	(4.29)	(18.80)
black def. × sex crime			-22.24**	33.83	-38.73***	1.56
			(9.46)	(31.60)	(11.22)	(27.81)
black def. × violent			12.78**	27.63	9.20	36.62
			(5.94)	(29.92)	(6.45)	(24.08)
black def. × drugs			-25.55***	-6.60	-23.26***	-10.72
			(6.79)	(20.22)	(5.68)	(20.39)
female def. × property			-9.71**	-22.80	-5.56	-35.49
			(4.81)	(17.74)	(5.26)	(22.64)
female def. × sex crime			-31.30**	21.98	-21.48	-64.30
			(15.23)	(80.65)	(15.92)	(57.63)
female def. × violent			-39.18***	-31.34	-34.05***	-77.26**
			(7.54)	(31.53)	(7.27)	(32.93)
female def. × drugs			-24.64***	-36.58*	-19.19***	-50.44
			(4.87)	(18.49)	(5.36)	(31.11)
Observations	76441	76441	76441	5681	61114	2758
Fixed Effects	No	Yes	Yes	Yes	Yes	Yes

Notes: Estimates denoted as including fixed effects are produced from models that include fixed effects for district and the rotation in which the offense was committed. Data range January 1, 2010, to December 31, 2011, because the rotation was interrupted in Fall 2009.

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Defendant characteristics are also important for determining the speed at which a case is disposed, which is evidence in Table 1.5. We show that female defendants take nearly 20 days longer to be added to the case management system and 40 days longer to adjudicate. In Table 1.6 we consider the effect of defendant gender and race as well as the interactions between those characteristics and broad crime types, and we consider the difference across characteristics *within* disposition types.

In columns 1-3, we consider all types of disposition outcomes together. In column 1, we show simple correlations supporting that female defendants' cases are resolved around 21 days more slowly than male defendants and that black defendants' cases are resolved 8 days faster. When we add crime type and district fixed effects to these OLS regressions in column 2, the results for black defendants become zero, but the effects for female defendants grow to around 23 days. Coefficients for all crime type categories are positive and statistically significant. The omitted category is a broad category of minor crimes, which are often handled more quickly. In column 3, we add separate interactions between crime type and whether the defendant is black or female. For black defendants, cases are slower for property and violent crimes, but faster for sex crimes and drug crimes. For female defendants, all interactions with crime type binary indicators are negative. Notably, the coefficient for whether a defendant is female increases to around 36.

We examine these relationships within types of dispositions. First, in column 4, we consider cases where all charges are dropped. Next, in column 5, we examine the effects on cases in which the defendant accepts a plea deal. Then, in column 6, we consider cases that result in a guilty verdict with no evidence of a plea deal. Consistently, we find that female defendants' cases are resolved more slowly, and that these effects are the most prominent for less serious crimes and not completely driven by whether the case ends in dismissal or a plea deal.

The relationship between defendant characteristics and case length is complicated by the fact that the way in which a case is resolved is also related to these two case attributes. Table 1.7 explores the relationship between case and defendant characteristics and the likelihood that all charges are dropped (columns 1-3) and that cases that are resolved through a plea deal (columns 4-6). In each set of columns, we first add fixed effects for district (columns 2 and 5) then we add controls and fixed effects for crime type and the rotation in which the crime occurred.

Importantly, female defendants are around 1.3 percentage points more likely to have their cases result in dismissal of all charges. If a case is heard by the third judge after the rotation in which a female defendant allegedly committed the crime in question, she is closer to 4 percentage points more likely to have all charges dismissed.[12]

[12]Female defendants' cases are dismissed around 9.7% of the time, so this effect is large - around 40% .

Table 1.7.
Defendant characteristics and rotations

	All Dismissed			Plea Deal		
female defendant	0.0124**	0.0137**	0.0130**	0.0072	0.0104	0.0038
	(0.0052)	(0.0063)	(0.0062)	(0.0084)	(0.0078)	(0.0074)
next judge	-0.0058	-0.0053	-0.0056	0.0781***	0.0712***	0.0708***
	(0.0037)	(0.0156)	(0.0159)	(0.0061)	(0.0233)	(0.0220)
3rd judge	0.0175***	0.0112	0.0110	0.0247***	0.0280	0.0396
	(0.0045)	(0.0272)	(0.0275)	(0.0070)	(0.0317)	(0.0304)
female def. × next judge	0.0137**	0.0117*	0.0115*	-0.0255**	-0.0287***	-0.0311***
	(0.0065)	(0.0067)	(0.0068)	(0.0101)	(0.0105)	(0.0099)
female def. × 3rd judge	0.0248***	0.0237**	0.0236**	-0.0300***	-0.0337***	-0.0408***
	(0.0079)	(0.0092)	(0.0091)	(0.0116)	(0.0101)	(0.0100)
black defendant	-0.0254***	-0.0098	-0.0107	0.0416***	0.0219*	0.0202*
	(0.0036)	(0.0072)	(0.0070)	(0.0063)	(0.0119)	(0.0111)
black def. × next judge	0.0076*	0.0016	0.0022	-0.0210***	-0.0184	-0.0139
	(0.0045)	(0.0085)	(0.0085)	(0.0076)	(0.0145)	(0.0137)
black def. × 3rd judge	0.0154***	0.0135	0.0147	-0.0230***	-0.0280	-0.0192
	(0.0056)	(0.0161)	(0.0161)	(0.0088)	(0.0190)	(0.0182)
Observations	66684	66684	66684	66684	66684	66684
Fixed Effects	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes

Notes: Estimates denoted as including fixed effects are produced from models that include fixed effects for district, the rotation in which the offense was committed and broad crime type. Data range January 1, 2010, to December 31, 2011, because the rotation was interrupted in Fall 2009. We also restrict to cases heard within 3 rotations of commit rotation (which is most cases).

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Results in this table also seem to suggest that female defendants are less likely to take plea deals under judges after the “current” period, and cases pertaining to black defendants are more likely to result in plea deals, especially in the “current” period. Despite this relationship between gender, dismissals and speed of adjudication, female defendants still have slower adjudication speeds than males. Column 4 of Table 1.6 shows that *within* cases in which all charges are dismissed, female defendants’ cases are still around 40 days slower on average. Also, for male defendants, around 20% of cases resulting in dismissal are disposed in the current period, while only 16% of similar cases for women are disposed in this period.

Next, we endeavor to determine whether the systematic delay of female defendants’ cases is driven by prosecutorial discretion. If it is, we expect to see evidence that the scheduling of female defendants’ cases is different from that of male defendants in a way that can be meaningful for outcomes. Table 1.8 contains evidence of this

Table 1.8.
Elected Judge Effect on Rotation

	All Defendants			Female Defendants		
	Current	Next	3rd	Current	Next	3rd
A. All Outcomes						
current judge elected	0.0013	-0.0110	0.0096	-0.0073	-0.0052	0.0125
	(0.0130)	(0.0184)	(0.0144)	(0.0213)	(0.0283)	(0.0256)
next judge elected	-0.0070	-0.0213	0.0283	0.0078	-0.0310	0.0232
	(0.0157)	(0.0244)	(0.0191)	(0.0244)	(0.0345)	(0.0329)
3rd judge elected	-0.0039	-0.0075	0.0114	0.0233	-0.0435	0.0202
	(0.0168)	(0.0224)	(0.0156)	(0.0210)	(0.0310)	(0.0223)
female defendant	-0.0248***	-0.0096	0.0345***			
	(0.0065)	(0.0064)	(0.0055)			
black defendant	-0.0080*	0.0011	0.0069	-0.0294***	0.0222**	0.0072
	(0.0045)	(0.0052)	(0.0052)	(0.0079)	(0.0099)	(0.0091)
Observations	66684	66684	66684	12007	12007	12007
B. Plea Deals						
current judge elected	0.0065	-0.0040	0.0051	0.0060	-0.0065	0.0012
	(0.0139)	(0.0159)	(0.0132)	(0.0190)	(0.0248)	(0.0210)
next judge elected	0.0050	-0.0259	0.0156	0.0303	-0.0435	-0.0080
	(0.0162)	(0.0213)	(0.0166)	(0.0230)	(0.0324)	(0.0268)
3rd judge elected	0.0076	-0.0088	0.0075	0.0484**	-0.0510*	-0.0011
	(0.0182)	(0.0188)	(0.0133)	(0.0215)	(0.0269)	(0.0198)
female defendant	-0.0204***	-0.0196***	0.0178***			
	(0.0050)	(0.0056)	(0.0052)			
black defendant	0.0001	-0.0007	0.0093*	-0.0220**	0.0124	0.0050
	(0.0041)	(0.0049)	(0.0054)	(0.0085)	(0.0097)	(0.0089)
Observations	66684	66684	66684	12007	12007	12007
C. All Dismissed						
current judge elected	0.0001	0.0004	0.0057*	-0.0071	0.0031	0.0141
	(0.0022)	(0.0034)	(0.0033)	(0.0050)	(0.0064)	(0.0086)
next judge elected	-0.0034	0.0020	0.0128***	-0.0117***	0.0117	0.0204**
	(0.0024)	(0.0045)	(0.0043)	(0.0043)	(0.0071)	(0.0088)
3rd judge elected	-0.0003	0.0072**	0.0049	-0.0083	0.0124**	0.0076
	(0.0024)	(0.0032)	(0.0034)	(0.0081)	(0.0059)	(0.0073)
female defendant	0.0014	0.0112***	0.0133***			
	(0.0017)	(0.0025)	(0.0028)			
black defendant	0.0006	-0.0034*	-0.0027	-0.0002	0.0041	0.0065
	(0.0016)	(0.0017)	(0.0018)	(0.0025)	(0.0029)	(0.0040)
Observations	66684	66684	66684	12007	12007	12007

Notes: Estimates denoted as including fixed effects are produced from models that include fixed effects for district, the rotation in which the offense was committed and broad crime type. Data range January 1, 2010, to December 31, 2011, because the rotation was interrupted in Fall 2009. We also restrict to cases heard within 3 rotations of commit rotation (which is most cases).

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

manipulation related to whether the presiding judge is locally elected at the time the case was disposed. In Panel A the outcome is whether the case was disposed in the current (column 1), next (column 2) or the 3rd rotation (column 3) from the rotation during which the crime was committed. In Panel B, the outcome is whether the defendant accepts a plea deal during the rotation for which the column is labeled. In Panel C, the outcome is similarly, whether all charges against the defendant are dropped in the relevant rotation.[13]

In addition to controlling for defendant characteristics, crime type and district fixed effects, we control for important temporal variables. We control for which "third" of the rotation the crime was committed and fixed effects for the rotation of crime commit date. The variables of interest pertain to whether the judges presiding in the current and upcoming rotations are locally-elected. Because most districts have more than one judge assigned at a time, we consider the proportion of judges that are locally-elected in each rotation. We focus on the current rotation, next rotation and 3rd rotation. Under normal circumstances, most cases will be disposed in one of these three rotations. Importantly, the characteristics of possible judges for each defendant depend on when the crime was committed. We believe that crime commit date is exogenous to judge characteristics, especially given that many cases are not heard by the judge who is currently presiding.

If prosecutors aim to match defendants to judges in a way that takes into account whether the judge is locally-elected, we could observe that these variables have an effect on whether the case is disposed in a given period. In fact, this is what we see in Panel C. If all of the judges in the 3rd rotation are locally-elected, then the defendant is more likely to have his or her case dismissed in the next (2nd) rotation. This effect is mirrored in the next column where it shows that if the next judge is locally-elected, then the case is more likely to be dismissed in the 3rd rotation.

Because female defendants exhibit the most interesting evidence of manipulation, we also replicate this analysis for female defendants. Columns 4-6 contain these results. Panel C displays results that suggest that the effects for dismissed cases described above are driven by female defendants. These results appear to suggest that the dismissals occur when there are no (or less) locally-elected judges presiding in a district.

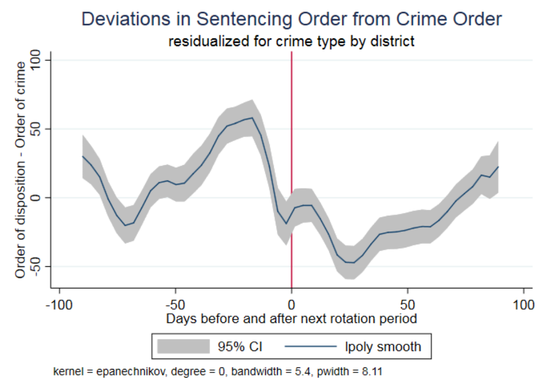
[13]Often, some charges are dropped by prosecutors, while others are carried through to adjudication. In this table, we focus on the cases where all charges are dropped and acknowledge that many plea deals result in some dropped charges.

Another way to consider this is looking at the types of cases, by timing, that are disposed around rotation periods changes. To investigate whether there is clumping of case timing decisions around rotation shifts we ran a kernel-weighted local polynomial fit of the residualized difference between disposition order and case order. Figure 1.4(a) presents this fit with a 95% confidence interval. The vertical line is at the first day a new rotation schedule goes into effect. As can be seen, there is an spike and sudden decline in 'older' (by order) cases right before a rotation change, followed by a period of 'younger' cases being disposed at the beginning of a new rotation. This could possibly be explained by prosecutors deciding to dismiss cases that are languishing or using the upcoming new judges as threat points to solicit plea deals from defendants.

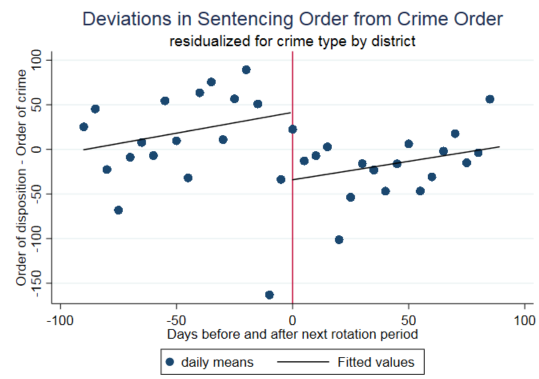
Since the kernel-weighted local polynomial may overfit noisy data, we also include a scatterplot of the daily means of the residuals both plotted against the days to (since) the next (last) rotation change. Figure 1.4(b) presents these results. While the intensity of the spike in 'older' cases is less pronounced, there remains noticeable differences between the age of cases resolved just before and just after a rotation shift. While the trend lines before and after a rotation shift are positive and similar, there is a downward shift in the average gap between disposition order and commit order at the beginning of a new period.

Whether this variation in timing is driven by strategic prosecutor decisions or not, these figures suggest that there may be differences in the types of cases that are disposed around rotations. If these are driven by judicial matching, the use of threat points in plea bargaining negotiations, or sorting on unobservables within and across rotation periods, then judicial assignment through rotation may fail to be exogenous if they are not adequately controlled for. It is also possible that this variation is driven by particular types of dispositions, for example dismissals, or by other institutional features, for example the impact of holidays on court schedules. Importantly, any of these scenarios can impact the sentences that defendants receive.

Table 1.9 presents results from regressions of sentencing variables (midpoint of total concurrent sentence and an indicator taking the value of 1 if a sentence was active) on these timing variables. Columns 1-3 present results for the midpoint total concurrent sentences and Columns 4-6 results for active sentences. There is a persistent and significant sentence penalty (longer sentence) for cases disposed in periods further away from the offense period. Interestingly, there is also a corresponding decrease in the extensive margin of sentencing into active incarceration as cases are disposed



(a) Kernel-smoothed fits of residualized differences in the order in which a case was disposed relative to the order in which the last corresponding crime in the case was committed. Order differences are residualized for district by crime type fixed effects. They are plotted against days since the scheduled set of judges changed. The vertical line denotes the first day new judges presided in a district.



(b) Daily means of residualized differences in the order in which a case was disposed relative to the order in which the last corresponding crime in the case was committed. Order differences are residualized for district by crime type fixed effects. They are plotted against days since the scheduled set of judges changed. The vertical line denotes the first day new judges presided in a district. Linear fits of the data are also displayed fit to either side of the judge change.

Figure 1.4. Plots of case disposition order against offense timing around rotation

further away from the offense period. These two results combined may partly reflect the fact that, when judges are able to choose between probation and active incarceration, they may be more likely to give higher sentences for probation, a pattern we do observe in this data and that the distribution of case types delayed into later periods is different than those heard in the current or next period. We observe a sentence penalty of around 41 days by judges in their election district when accounting for both rotation window and offense period for cases ruled, although this result change signs in the instrumental variables models.

Sentences of offenses that were committed in the first third of a period are roughly 2.5 days longer on average than sentences for charges committed in the last third of a period, but this result is not significant. This compares sentences regardless of the period in which they were disposed in. Since offenses committed in the first and last thirds of a period are likely to be disposed during different rotations, we interact the first-third indicator with indicators for each rotation window. Offenses committed in the first third of a period have sentences roughly 48 days longer on average when disposed in the same period as the offense, but by the third and fourth period cases

Table 1.9.
Timing variables on sentence variables.

	Total concurrent sentence (midpoint)			Active		
Rotation period=2	73.46***		82.89***	-0.0291***		-0.0434***
	(9.677)		(9.634)	(0.00736)		(0.0111)
Rotation period=3	111.1***		144.0***	-0.0520***		-0.0693***
	(13.81)		(15.47)	(0.0107)		(0.0140)
Rotation period=4	115.6***		150.7***	-0.0581***		-0.0831***
	(13.92)		(14.79)	(0.0110)		(0.0147)
Offense in first third of period		2.486	49.08***		-0.000616	-0.0266**
		(1.695)	(5.828)		(0.00315)	(0.0102)
Rotation period=2 × Offense in first third of period			2.247			0.00951
			(8.360)			(0.0133)
Rotation period=3 × Offense in first third of period			-41.92***			0.0103
			(8.464)			(0.0110)
Rotation period=4 × Offense in first third of period			-51.79***			0.0248*
			(10.46)			(0.0125)
Election district	41.17***	48.47***	41.37***	0.0208**	0.0206**	0.0234**
	(10.72)	(11.97)	(10.37)	(0.00867)	(0.00852)	(0.00890)
Plea	58.48***	61.36***	60.75***	0.0213*	0.0220*	0.0192*
	(6.131)	(7.469)	(6.281)	(0.0109)	(0.0115)	(0.0109)
jury trial	398.5***	425.8***	390.1***	0.325***	0.306***	0.323***
	(31.71)	(33.48)	(31.95)	(0.0148)	(0.0211)	(0.0209)
Observations	200751	133564	133564	172422	114819	114819
District FE	Y	Y	Y	Y	Y	Y
Crime type FE	Y	Y	Y	Y	Y	Y
Case controls	Y	Y	Y	Y	Y	Y

Notes: Standard errors clustered on the district-level are reported in parentheses. Results are obtained from regressions of Equation 1.6 and Equation 1.7. All models include crime type fixed effects. Columns 2, 3, 5 and 6 compare offenses committed in the first third of a period to offenses committed in the last third of a period.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

with offense commit dates in the last third of a period have higher sentences by around 40-44 days on average. There are small and insignificant differences between cases committed in the first third and last third of a period when disposed in the immediately proceeding rotation period.

At the extensive margin, later periods continue to have a declining proportion of active sentences. As can be seen in Column 6, offenses committed in the first third have 2.7 percentage points less active sentences than those committed in the last third when disposed in the same period. For cases disposed further away than the offense period, there are small and insignificant differences between the first third and last third groups until the fourth rotation window, with more active sentences observed for offenses from the first third. Similar to above, this diverging share of active sentences combined with changes to sentencing patterns suggest that the development of cases over time may result in different distributions of case types reaching judges over time, depending on when the offense was committed in a rotation period.

1.6 Discussion

In this chapter, we consider the possibility the prosecutors match defendants and cases to specific judges when provided a situation that allows for such strategic behavior. We focus on North Carolina where judges are elected to serve a specific geographic area, and then they only spend some of their time presiding in their own constituency.

Results point to an important indication that there are systematic relationships between the timing of a case, the set of rotation periods that a case can be resolved in, and the case disposition. When considering average sentences and the extensive margin of sentencing into active incarceration, the distribution of case outcomes shifts over time depending on whether the offense was committed in the first third or last third of a rotation window. The timing of an offense in the first third, conditional on the duration from offense to involvement of a prosecutor, significantly opens up the possibility of resolving a case in the current period while retaining some options to delay a case to future periods. When comparing offense timing further into the future, there is also a sentence penalty for offenses committed in the last third of a period, which may be easier to delay into later rotations without violating statutory case process requirements or increasing the likelihood the defense or judge asks for an speedier resolution.

Locally-elected judges appear to behave differently from non-local judges in our analysis. First, we show that locally-elected judges pass down harsher sentences. The persistent sentence penalty that comes from having a judge in their own election district, which may be driven by a judge's election concerns if their rulings are more exposed to the electorate, further underscore the importance prosecutors may place in considering which judge will be available. Second, we show that dismissals, particularly for female defendants, are less likely to occur when locally-elected judges are presiding in a district. These two effects could be related, For example, if judges are concerned about the optics of their rulings, they may want to be seen as "tough on crime," doling out harsh sentences and preferring not to be the judge of record on dismissals. Last, we show that when we are able to isolate the *causal* effects of having a locally-elected judge, they actually lead to a lower likelihood of taking a plea deal and shorter sentences.

In addition to evidence of sentencing effects from case timing, there is a noticeable shift in the types of cases, by timing, directly before and after a rotation changes. This result can be driven both by normal seasonal variation that happens to coincide with rotation shifts or by institutional features other than strategic behavior by prosecutors. That said, given the wide latitude prosecutors have in North Carolina to schedule cases, it is also possible these results represent strategic behavior in case scheduling to take advantage of shifting judicial contexts across rotations. All of these results together suggest that elected judges are sensitive to the concerns of their potential voters.

Importantly, if there is any strategic behavior occurring on the part of the prosecutors that also impacts sentencing or other outcomes, empirical approaches that rely on the judicial rotation calendar for random judge assignment or for case ordering may be threatened. Careful examination of the types of cases assigned to different judges reveals sufficient evidence to suggest that there is some sorting between judges, and empirical studies intending to use random judge assignment as an instrument should be aware of this possibility.

1.7 Appendix

1.7.1 Data cleaning

In order to complete the replication of Abrams et al. (2018), we follow the data cleaning process found in the Appendix of Abrams and Fackler (2018), with the following exceptions.

We use data from the North Carolina Department of Public Safety Offender Public Information database online to collect all known aliases of defendants. Aliases include alternative names, but also alternative spellings, inclusion of suffixes or hyphens, or other transcription errors that are found in the ACIS system but associated with a unique offender. We assign a single alias to all names in the DPS Public Information File and then match by name and birthday with the original ACIS data. Finally we replace all names that have a known alias with the alias before grouping together in cases. This prevents alternative spellings or known transcription errors from being treated as separate individuals.

For much of our analysis we need to identify a single offense commit date for a case. Since cases often combine multiple offenses that may have different commit dates, we choose the offense date that happened closet to the case disposition date (as defined by Abrams and Fackler (2018) but not after the case disposition date. Since the case disposition date avoids assigning a disposition date associated with a parole or probation revocation to the case unless the original case can not be identified, this prevents using offense dates for parole or probation violations. We chose the date closet to disposition as we believed that it is the offense most likely to be close to when the prosecutor becomes involved in the case. We also ran a robustness check on all tables and figured that depended on the case offense date by using the *earliest* offense date before case disposition. We do not find any change to the practical results. This is likely because most cases involve offenses committed within a similar time frame, with the exceptions being extreme results in both analyses. Robustness results from this alternative case date specification are available on request. We also identify a single case system create date for each case, which Abrams and Fackler (2018) do not discuss. For this we chose the earliest system create date for all charges with the case offense date.

In addition to the set of case types omitted by Abrams and Fackler (2018) and discussed above, we also dropped cases that carried with them the possibility of life without parole that were not already excluded by dropping homicides and violent sex crimes. We also excluded DWIs, since these are sentenced outside of structured sentencing and have a wide variation in sentencing for similar charged offenses (DWIs are typically charged with the same offense code, but then convicted on an offense code that takes into account multiple aggravating factors).

CHAPTER 2. A REVIEW AND COMPARISON OF JUDICIAL SEVERITY INDEX CALCULATIONS

2.1 Introduction

In situations where outcomes depend on the decision-making of a third party, such as sentencing by judges in criminal courts, rulings on disability insurance cases by administrative judges, the results of arbitration by mediators, grading of assignments by teaching assistants, or scoring of business investment plans by analysts, the propensity towards particular types of decisions by the third-party may have significant effects on the outcomes that are credibly exogenous to other observable and unobservable determinants of the outcome. If the third-party decision maker is as-good-as-randomly assigned, these differences in propensities can be used as instruments for the likely endogenous decision itself. A classic example is a judicial severity index, which measures the propensity of randomly assigned criminal court judges to sentence or incarcerate and applies this measure as an instrument for the sentence or incarceration spell itself.

As a result, measuring the propensity of third-party decision makers is a core component of the toolkit of the applied economist. It has become particularly important in the use of so-called "random assignment of judges" in the economics of crime (Kling, 2006; Dahl et al., 2014; Dobbie et al., 2018) and in the analysis of administrative court rulings on labor-market or welfare outcomes (Maestas et al., 2013; Dobbie and Song, 2015). The generalized methodology has also been employed outside the scope of judge assignment to a wider range of situations that depend on third-party decision makers with differing propensities to make assignment decisions (Jr. et al., 2015; Gonzalez-Uribe and Reyes, 2019). It is easy to imagine how the methodology can be employed not only in the setting of criminal justice or administrative courts, but also in scoring of investment or grant proposals, grading of student work, or in matching problems that involve some form of human scoring.

The application of a propensity index as an instrumental variable comes with the typical requirements for relevance and validity of the instrument. In addition, inference of local average treatment effects requires the index to be well-behaved, which most commonly means that it be monotonic (see below for a more detailed discussion of monotonicity in this context). While a body of literature has developed methods for assessing the validity and behavior of severity indexes in general (Frandsen et al., 2019; Norris, 2019), it is difficult to test for violations of the exclusion restriction. Frandsen et al. (2019) propose a joint test for monotonicity and the exclusion restriction, but cannot tell the two apart. They propose placing greater weight on failure from violations of monotonicity, but it is not clear under what conditions this result will hold. This is particularly true if judges are likely to influence case outcomes through multiple channels, for example through the decision of whether or not to incarcerate and through the decision on how long the sentence should be, both independently from the decision to incarcerate and conditional on it. These multiple channels may be adequately captured by a single severity index (for example one that measures the propensity to incarcerate), but we are unaware of any work that has tested this assumption.

Assessing possible failure of the exclusion restriction is complicated by the fact that applications of judicial severity indexes have focused on the use of a single index measure. In practice, one can imagine a wide range of different possible index calculations that may reflect the propensities of assigned judges—for example the propensity to incarcerate, the propensity to assign probation, the propensity to sentence relative to sentencing guidelines, or the propensity to sentence conditioned on the propensity to incarcerate or assign to probation[1]. If all possible propensity indexes are relevant, valid, and well-behaved, we expect results to fall within a similar range of outcomes. However it is possible that judges differ in their propensity based on differing measures, or that differing measures capture slightly different degrees of judicial influence on outcomes. While we cannot conclusively determine that a measure is valid and well-behaved merely by comparing it to alternative specifications, we can highlight possible issues that can be addressed by more nuanced measures of judge severity.

This work considers this issue by constructing a wide range of severity indexes for the same administrative dataset from the universe of cases heard by the Superior Court of North Carolina from 2008-2012. Indexes are calculated that measure propensities to incarcerate, as well as propensity to sentence to two different types of

[1]As discussed on the third season of the NPR podcast *Serial* in the case of Cleveland, it may be that some judges prefer to assign defendants to probation with long sentences, especially if they prefer to have repeated supervision over a defendants rehabilitation or if they want to ensure that they will be responsible for sentencing a defendant to active incarceration if and when they recidivate. In this case, a judge may appear to have a lower propensity to incarcerate (i.e. appear ‘less’ severe on an incarceration measure) but have a propensity to longer sentences relative to sentencing guidelines (i.e. appear ‘more’ severe on a sentencing measure)

probation that differ in supervision level. Additionally, sentencing relative to possible maximum and minimum sentences given by North Carolina's structured sentencing programs are used to construct additional measures of judicial severity that will help capture a wider range of case outcomes based on the range of charges and the prior record of defendants. Finally, an index that measures the propensity to approve requests for continuations is constructed as a test of whether judge leniency or judge type can also be observed from procedural decisions. Data on cases disposed from 2008-2012 is then combined with data from the North Carolina Department of Corrections on cases from 2013-2019 to construct a measure of recidivism for defendants in the original dataset. Each index is then used in two-stage-least-squares to instrument for sentence type (incarceration, intermediate probation, or community probation) and sentence duration (both total sentence and total duration of incarceration).

Finally, judges are classified as types based on their combined propensities to incarcerate or sentence to community probation, the least supervised sentence type. The effect of assignment to judge types on recidivism directly are analyzed, and the typology is combined with the instrument approach for instrumenting assignment given by Carr and McClain (2019). As judges rotate throughout districts in North Carolina, defendants are exposed to different possible sets of judicial assignments based on previously released schedules. Assuming some control over scheduling by prosecutors, we instrument for assignment using the week of offense commit date to predict the probability a case is disposed in the same or next period after the offense was committed. We examine whether there are systematic deviations in the likelihood of having judge types by defendant characteristics.

In general, we find that the majority of our severity indexes pass the weighted average monotonicty tests proposed by Norris et al. (2019), although some specifications fail to find significant results for female defendants. Likewise, balance tests are generally successful for most of the primary case and defendant characteristics used in these tests for propensity to incarcerate. However there are several notable failures, especially for propensity to sentence to probation and propensity to sentence intensity (actual sentence relative to possible sentence). These failures center primarily on case types, female defendants and judges in their home district. This confirms similar observations of possible strategic assignment of cases around home judges and female defendants found by Carr and McClain (2019). Additionally, there appears to be a systematic tendency for jury trials to be heard by more severe judges, a case characteristic often excluded in balance tests.

Point estimates from two-stage least-squares estimates are broadly consistent across instruments, whether considering the effect of incarceration, probation, or sentence duration. The notable exception are statistically significant differences between point estimates when using sentence intensity as an instrument and conditioning on whether the sentence was probation or incarceration, although the sign of these effects are the same. We believe this highlights the fact that judicial propensities may be multidimensional, and that judges may behave differently when they are sentencing defendants to probation and when they are sentencing to active incarceration. Additionally, we document that only community probation leads to lower recidivism rates overall. The result is mirrored when we consider judge types: defendants assigned judges who are more likely than their peers to sentence defendants to community probation and less likely than their peers to sentence defendants to incarceration, a judge type we label "rehabbers," are 0.5% less likely to re-offend. Conversely, defendants assigned to judges more likely than their peers to incarcerate and less likely to sentence to community probation, a judge type we label "harsh," are 0.5% more likely to re-offend. In addition, black defendants are 1% more likely to be assigned to harsh judges, but 1% less likely to be assigned to judges sentencing to community probation. The exact opposite is true for female defendants: they are 1.3% less likely to be assigned to harsh judges and 1.4% more likely to be assigned to judges sentencing community probation.

2.2 Use and conditions of severity indexes

Severity indexes have been a mainstay of the economics of crime literature with the increased use of random assignment of judges as an instrument for incarceration, pre-trial detention, or sentences. An instructive example is given by Kling (2006), who uses random judge assignment and sentencing propensities of judges in Florida and California to evaluate the effect of incarceration length on future employment and earnings. It is clear that sentence length will be driven by offense severity and criminal history, both of which will likewise be correlated with unobservable defendant characteristics that will also influence future employment and earnings. To estimate the causal effect of an additional year of incarceration, this sample selection problem would have to be dealt with. Kling addresses this issue by employing judge fixed-effects to estimate propensities to sentence. These propensities are then used as an instrument in a two-stage least squares estimation procedure to estimate the causal effect of incarceration length on employment and earnings. In effect, Kling estimates the effect of being assigned to a more or less severe judge on similar groups of defendants, so that differences in sentence lengths are driven by the exogenous and random assignment of a judge rather than the (endogenous) unobservables that drive both offense type, criminal record, defendant behavior on trial, and labor market outcomes.

The method has been widely applied by researchers in the economics of crime. Dobbie et al. (2018), for example, use random assignment and propensities to detain defendants pretrial to estimate the effect of defendants being detained prior to their trial on convictions, future crime, and employment. Norris et al. (2019) use random assignment of judges in Cleveland, Ohio to estimate propensities to incarcerate, which they use to estimate the effect of parental or sibling incarceration on academic, socioeconomic, and criminal justice outcomes of children. The methodology has also been extended to administrative court settings (Maestas et al., 2013; Dobbie and Song, 2015).

The basic methodology is largely preserved in all of these papers: a severity/propensity index is constructed by comparing the individual effects of assigned judges to area and time means, typically through the use of judge fixed effects on observations of sentences, incarceration, or other judicial decisions residualized by location and time. The validity of a severity index requires that the propensity of the judge be independent from the unobserved differences in the samples of defendants heard by each judge (i.e. the judge is randomly assigned) and that the propensity measured by the index is the only channel by which the assigned judge impacts outcomes (i.e. the exclusion restriction, that judges do not influence future outcomes except by the decision the propensity instruments for).

The core assumption necessary for judge fixed effects to be valid is that of random assignment, or the exclusion restriction. This condition is often tested simply through balance tests on observables, which may fail to detect assignment on unobservables or non-random assignment for some cases that, on average, result in apparently balanced samples. Since these efforts may only shift likelihoods of assignment and depend on the schedule of available judges, balance tests alone may fail to detect underlying shifts in the likelihood of assignment, especially if they do not condition on changes to judicial availability. In addition, the threat of strategic judge assignment may be used by prosecutors as a threat point in plea bargaining negotiations. If these threats are successful, the assigned judge may not be the judge relevant for influencing the plea bargain sentence, potentially threatening the reliability of the index itself and failing to be detected through balance tests.

In addition to the exclusion restriction, inference of a weighted average of treatment effects from the use of a severity index requires monotonicity of the instrument, which implies that an increase in the severity index will represent an increase in the actual realization of that propensity. In other words, a strict form of pairwise monotonicity implies that an individual sentenced by a less severe judge must be sentenced by a more severe judge. Frandsen et al. (2019) and Norris (2019) suggest alternative monotonicity restricts which depend only on the average effect of increasing severity. Average monotonicity implies only that the covariance between the severity index and the decision being instrumented for (e.g. incarceration, sentence length) be the same sign for compliers within the dataset (Norris et al., 2019). Dobbie et al. (2018) employ a test that can reveal meaningful information on weighted average monotonicity that involves running first-stage regressions on different sub-samples and ensuring that the instrument has the same sign for all sub-samples.

Frandsen et al. (2019) propose a joint test for exclusion and pairwise monotonicity based on the fact that a well-behaved and valid index will lead to marginal treatment effects bounded by the support of the outcome variable, but this test cannot differentiate between violations of monotonicity and the exclusion restriction. Their test first regresses the outcome of interest (for example binary indicators for recidivism or labor force participation) on a flexible function of the propensity index, proposing a non-parametric b-spline or local polynomial for the flexible function. The residuals from this regression are then regressed on judge indicators and testing whether the coefficients on judge indicators are jointly zero. They also test whether the slopes of this flexible function are within the bounds implied by the support of the outcome variable. Frandsen et al. (2019) suggest that, under prior assumptions of validity of the exclusion restriction, a greater weight should be placed on the likelihood that test failures are violations of monotonicity and propose a weaker condition of weighted average monotonicity under which IV estimates are asymptotically valid.

While under ideal conditions it may be reasonable to assume the exclusion restriction holds, it is not immediately clear that this is always the correct assumption. The very concept of judicial severity indexes—that judges have differing propensities to incarcerate or sentence—belies the notion of random assignment if the prosecutor or defense have any ability to influence assignment, for example through strategic scheduling decisions by the prosecutor or strategic motions by the defense attorney

(Carr and McClain, 2019). If judges truly do have differing propensities that will have significant effects on the likely outcome of a case, it should not be surprising if there are efforts at influencing assignment. In addition, it seems unlikely that the only effect judges have on outcomes are through the decision to detain or incarcerate. Judges will rule on motions concerning warrants, dismissals, continuation or evidence. Even in the absence of these motions, they will likely play a significant role in the management of a case through the system by approving hearing dates or plea bargains.

This work adds to the understanding of the exclusion restriction on severity indexes by comparing results from multiple alternative specifications of judicial propensities. Using data that allows for calculation of propensities to incarcerate, to sentence to two different types of probation, to set sentence lengths relative to structured sentencing mandates, and to rule on motions for continuance, we construct a wide range of alternative judicial propensity indexes for the same set of judges in North Carolina. Using data on criminal convictions following our original data time horizon, we compare the results from this broad array of different severity indexes at satisfying conditions for validity and montoncity and estimating two-stage least squares estimates of incarceration and relative sentence length on recidivism.

We expect that under the exclusion restriction, any single valid, relevant and well-behaved index should provide an estimate of the weighted average treatment effects consistent with those from other valid, relevant and well-behaved indexes. While observing consistent results across specifications is not sufficient for concluding that all indexes are valid and well-behaved (especially if all indexes suffer from the same violations), a review of multiple competing methods for estimating judicial propensities can aid in supporting additional arguments and test for valid and well-behaved instruments. The opposite also holds: while failure to replicate results using different indexes does not provide clear evidence of which (if any) index is valid and well-behaved, it can direct the researcher towards possible threats to validity or monotonicity violations.

2.3 Data and institutional setting

Data come from the 2012 Criminal Case Information Statistical Extract from the North Carolina Administrative Office of the Courts. Data are collected from the Automated Criminal/Infraction System (ACIS), which includes administrative data recorded by clerks in the Superior Court of North Carolina, and are available for all criminal charges originating in the state of North Carolina, whether they led to a conviction or not. We use the same data cleaning and observation identification process as Carr and McClain (2019), resulting in a final dataset of 236,963 cases consisting of 1,563,999 charges (including charges that were dismissed or superseded). Table 1.1 in Chapter 1 provides descriptive statistics on the dataset, which is the same as employed in that chapter.

The data include information on the date of the charged offense, the zip code, race and sex of the defendant, the specific crime that was committed, the prosecuting attorney, the assigned judge, the disposition method, the plea entered by the defendant, and additional dates and indicators of case development. Adjudicated charges also include maximum and minimum sentences, sentence type, probation length, and fines. We combine this with data on the structured sentencing system in place over our data's time period. In North Carolina, a grid is developed that, for each charge class[2] and prior record points[3], a range of possible minimum sentences are given at the sentencing judge's discretion. Each possible minimum sentence also carries with a corresponding maximum sentence, given in an additional document. This allows us to calculate severity indexes that compare the actual minimum (or maximum) sentence with the range of possible minimum (or maximum) sentences that can be given for the set of charges a defendant originally faced. In addition, each cell on the structured sentencing grid includes whether the judge may assign active incarceration, intermediate probation (a form of probation that often involves some supervised detention or stringent conditions), or community probation (a probation with less intense supervision and some additional requirements, e.g. substance abuse treatment). This also allows us to condition probabilities of incarceration or probation on whether a incarceration or probation was a possible outcome for a given case. More information on how we use the structured sentencing grids is given below. During our time period, three different structured sentencing grids were in effect. New grids went into effect on 12/1/2009 and 12/1/2011. The December 2009 changes shifted

[2]There are ten charge classes for felonies, given as letters from A to I (with two levels at class B) that range from most serious (A) to least serious (I). There are four classes for misdemeanors (A1, 1, 2, and 3), with A1 being most severe and 3 being least severe. Our dataset focuses on felony cases.

[3]Offenders are given record points following convictions. Points vary based on the class of the charge, with Class A felonies worth 10 points and Class H or I felonies worth 2 points. Additional factors, such as parole/probation violations or crimes committed while incarcerated, may also result in additional points. The structured sentencing grids break prior records into six ranges for felonies, which change slightly over our time period.

the ranges of prior record points and had some modifications to possible minimum sentences and the December 2011 changes increased the range of possible minimum sentences for offense classes F through I. In addition, some grid cells had changes to whether active incarceration or probation were options.

The use of severity indexes as an instrument for judges depends on random assignment, which can be satisfied in North Carolina through the rotation of judges across districts. North Carolina court system has 3 levels: counties, districts and divisions. The state is divided into 8 divisions, with 5 to 7 districts within each division for a total of 50 districts. With the exception of the most populated urban areas, counties are completely contained within a district. For less populated areas, it is possible that the same district will contain more than one county. Superior Court judges are elected on the district-level to eight-year terms, providing them with a "home" district and division. Following retirement or other causes of a judge leaving the bench prior to the end of their term, the remaining years are served through appointment by the governor.

A "rotation" in North Carolina represents a six-month period where each judge is assigned to a district, with rotations beginning on either January 1 (spring rotation) or July 1 (fall rotation). Judges are guaranteed to have one rotation period a year in their election district and will serve the other six months in another district within their election division. The governor has the authority to temporarily cancel the judicial rotation schedule, effectively returning all judges to their home district. This happened once in our time frame, in July and August 2009[4]. The average number of judges in a district is just over two, with districts ranging from having only one judge to seven.

The Chief Justice of the North Carolina Supreme Court sets the calendars, often with input from division chiefs. Individual judges do have the ability to request changes to the schedule, although this does not appear to be a widespread practice. The assistant director of the Administrative Office of the Courts assists the Chief Justice in scheduling and also manages changes to the master calendar due to sickness, recusal, or other legitimate reasons for a judge being unable to meet their schedule requirements. We make use of these schedules in the last portion of this paper, where we combine severity indexes with an instrumental variables approach that attempts to control for the probability a specific set of judges may hear a case. More on this is discussed below.

[4]In all subsequent regressions, we exclude data from the rotation period with the stoppage.

2.3.1 Recidivism data

For our two-stage least squares estimation, we collect data on recidivism from the North Carolina Department of Public Safety (DPS) offender database. This publicly available database includes all individuals who have been incarcerated or placed on probation in the state of North Carolina, including data on their offense and incarceration dates. We collect data from the DPS Offender Public Information for offenses committed in the period of 2013 to 2019, so there is no overlap with offenses that would have been committed in our original data. We then match observations by name and zip code to our original data on criminal charges. All matched individuals represent individuals convicted of crimes committed after 2012, who were charged with a crime in between 2009 and 2012. In other words, everyone matched between these two sets of data committed an additional crime and is considered a recidivist. We also observe the number of additional charges committed after 2012 for all matched individuals.

There are three primary issues with measuring recidivism this way. First, recidivism does not require that a re-offending criminal is caught. As a result, any data on recidivism that is based on criminal justice system data naturally excludes a portion of re-offenders who have not been identified by police prior to the use of the data. This is a common issue for any paper studying recidivism. As a result, measures of recidivism from police, court or corrections records should be seen as a lower bound on actual recidivism.

A second, and related, issue is that not all re-offending criminals will be incarcerated or sentenced to probation. Incarceration depends not only on apprehension, but also process through the court system and eventual sentencing to prison or probation. Since our measure of recidivism involves matching offenders to data on these measures, we likewise may miss some additional re-offenders who were fined or had their charges dismissed, even if they have been apprehended within our time frame. In addition, incarceration following recidivism is likely to be significantly higher (and more likely to be observed within our time frame) if the offender had initially been sentence to supervised probation, a suspended sentence, or was already incarcerated. All of these features bias our estimation of recidivism based on actual sentencing to probation upwards.

Finally, recidivism is a lifelong issue. We only observe a limited range of time following the end of our initial time frame, and so it is possible that there are offenders who have not re-offended but will again in the future. That said, the duration between release and re-offense is often quite short (Uggen, 2000). A study that considered release of prisoners in 2005 found that over two-thirds of re-offenders are arrested within three years, and more than three-quarters are arrested within five years. Our time frame considers incarceration within seven years of the latest period in our initial sample, and so covers the most likely period of re-offense.

2.4 Empirical strategy

2.4.1 Calculating measures of sentence type and intensity

We construct a broad set of measures, M^n of each type, for measuring propensity to sentence to a particular type of sentence (incarceration or probation), propensity to sentence to longer duration relative to possible sentence lengths (sentence intensity), and propensity to approve continuations in a trial. Results on the full range of measures employed is included in the appendix, but for our main results we will focus on a subset of measures and propensity calculations. Table 2.1 gives descriptions of the eight measures employed in the main body of this work[5]. Measures are constructed from observational data from North Carolina, with some transformations for sentence intensity and one sentence type measure.

Measures for sentence type are observed as a binary indicator for whether any of the sentences given for a case were active incarceration or probation (either intermediate or community), such that $M_i^{type,m} \in \{0,1\}$ where $m \in \{active,\ intermediate,\ community\}$. Since cases can have multiple sentence types if defendants are convicted for more than one charge, some of which carry active incarceration and some of which carry probation, we also construct a categorical variable that indicates the worst sentence type given in a case:

$$M_i^{type,worst} = \begin{cases} 1 & \text{if no sentence} \\ 2 & \text{if worst sentence is community probation} \\ 3 & \text{if worst sentence is intermediate probation} \\ 4 & \text{if worst sentence is active incarceration.} \end{cases} \tag{2.1}$$

[5]Table 2.16 contains means, standard deviations, minimums and maximums of the complete set of measures used to construct propensity indexes beyond the primary eight used in the main body of this work.

Table 2.1.
Descriptions of primary measures used in propensity index construction

Measure		Description	Mean (Std. Dev.)
Active Incarceration	$M^{type,active}$	An indicator variable equal to 1 if *any* of the charges given were active incarceration	0.27 (0.45)
Intermediate Probation	$M^{type,inter}$	An indicator variable equal to 1 if *any* of the charges given were intermediate probation	0.27 (0.44)
Community Probation	$M^{type,comm}$	An indicator variable equal to 1 if *any* of the charges given were community probation	0.35 (0.48)
Worst outcome	$M^{type,worst}$	A categorical variable that measures the highest sentence type, from no sentence (1) to active incarceration (4)	2.65 (1.04)
Midpoint to case worst max minimum	$M^{intensity,wmax}$	The ratio of the midpoint of a given sentence to the highest maximum minimum possible for all charges in a case, regardless of whether they were sentenced	0.49 (1.25)
Midpoint to case best max minimum	$M^{intensity,bmax}$	The ratio of the midpoint of a given sentence to the lowest maximum minimum possible for all charges in a case, regardless of whether they were sentenced	2.55 (10.35)
Midpoint to sentenced charges worst max minimum	$M^{intensity,max}$	The ratio of the midpoint of a given sentence to the highest maximum minimum possible for all charges sentenced	0.62 (1.05)
Midpoint to case worst min minimum	$M^{intensity,wmin}$	The ratio of the midpoint of a given sentence to the highest maximum minimum possible for all charges in a case, regardless of whether they were sentenced	2.07 (43.15)
Total Continuations Granted	$M^{motions}$	The total number of continuations approved by a judge	0.90 (1.61)

For measures of sentence intensity, we make use of the structured sentencing grid that gives ranges of minimum sentences that may be applied given a charge class and defendant prior record points. There are multiple ways we can consider sentence intensity. Comparing the midpoint of the sentence range or the minimum sentence given to the minimum or maximum minimum sentence possible at the case level[6] (i.e. if the only sentence was for the charge with the lowest or highest possible sentence) and comparing the midpoint of the sentence range or the minimum sentence given to the minimum or maximum minimum sentence possible at the charge level (i.e. directly comparing the sentence given for a charge to the range of possible sentences for that charge) are the two predominant measures employed in this work. Sentence intensity is calculated as the ratio of the observed sentence S_i to a range of different sentences that could have been given based on the set of charges in a case and the structured sentencing grid, $\overline{S_i^n}$:

$$M_i^{intensity,n} = \frac{S_i}{\overline{S_i^n}}. \tag{2.2}$$

$M^{intensity}$ measures are thus a real number with bounds depending on the possible sentence being applied. For example, comparing the minimum sentence given to the smallest possible sentence at the case level will be bounded below at 0 (for no sentence) and above by the ratio of the highest possible sentence in the structured sentencing grid to the lowest possible sentence in the sentencing grid. For cases whose sentence is on the worst possible charge, the minimum sentence given to the maximum minimum sentence is bounded below at 0 and above at 1[7]. Finally, we construct a measure

[6]The structured sentencing grids give ranges of *minimums*, with each minimum having a corresponding maximum. We simply employ the structured sentencing grid minimums for all sentence intensity comparisons. Ranges between minimums and maximums vary only across two broad groups of offense classes (all offense classes B1 through E have the same minimum-maximum ranges and all offense classes F through I have the same ranges). So, for example, a minimum sentence of 25 months is possible for charges in offense classes H, G, F and E. For charges in offense classes H, G, and F, a minimum sentence of 25 months has a corresponding maximum of 39 months. For charges in offense class E, 25 months carries a maximum of 42 months. The lowest possible sentence for Class E is 15 months (mitigated with no prior record points). The highest possible sentence for Class F is 41 months (aggravated with > 18 points). We control for broad crime type and run robustness checks including offense class at various points of our analysis, and do not believe using only minimums in our measures of sentence intensity cause any systematic bias in our severity indexes.

[7]For multiple reasons, including measurement error, missing structured sentencing data on 'free text' charges, and the fact that we do not observe whether a sentence was in the mitigated or aggravated range and only use the presumptive range minimums in our intensity measures, it is possible that we observe values greater than 1 in this measure. We run multiple robustness checks taking this into account, including dropping all cases with missing structured sentencing data and winsorizing our measures prior to index construction.

for how judges rule on motions, $M_i^{motions}$, which is the total number of continuations approved by the judge for each case i and is a non-negative integer[8].

2.4.2 Calculating judicial severity indexes

We calculate two primary sets of judicial severity indexes that consider propensities to sentence types (active incarceration, intermediate probation, and community probation) and to sentence intensity (comparing the minimum sentence assigned in a case to different possible sentences given North Carolina's structured sentencing guidelines) and one additional index that measures the propensity for judges to approve motions for continuation. Calculating all of these propensities involves comparing judge effects to district-month means on a set of different measures, M_i^n, where M is the measure of propensity type n (where $n \in \{sentence\ type,\ sentence\ intensity,\ motions\}$) for case i. The primary method, which we call the *residuals method*, for calculating a propensity index of type n for judge j, p_j^n, is as follows:

1. Calculate measure M^n
2. Regress M^n on district-by-case month indicators
3. Calculate the residuals from the district-by-case month means, $\widehat{\varepsilon^n} = M - \widehat{M^n}$
4. Regress $\widehat{\varepsilon^n}$ on judge fixed effects
5. Calculate the severity index as the fitted values of this regression, $p_j^n = \widehat{\widehat{\varepsilon_j^n}}$.

In addition to the residuals method, we also employ what we call the *comparison method*:

1. Calculate measure M^n
2. Regress M^n on dummies for judges and district-case month indicators
3. Calculate the fitted values of this regression, $\widehat{M_{full}^n}$
4. Regress M^n on district-case month indicators alone

[8]It is possible that some continuations are approved by judges other than the sentencing judge. Since judges are set following the initial hearing scheduling, it is assumed that the majority of continuations are requested following the initial setting of a trial date and are likely to be the eventual sentencing judge. To the extent this is not true, our measure overestimates the tendency for a judge to approve continuation motions.

5. Calculate the fitted values of this regression, $\widehat{M^n_{district-month}}$

6. Calculate the severity index as the difference between the fitted values of the first regression and the fitted values of the second regression, $p^n_j = \widehat{M^n_{j,full}} - \widehat{M^n_{j,district-month}}$.

We consider the comparison method because it has the added benefit of judges having district-case month specific severity indexes rather than one severity index for all cases. In effect, using the comparison method we observe the propensity of a judge at a specific place and time, as opposed to the average propensity of a judge across all districts and case months. This allows for the possibility that a judges propensity develops as they gain experience in a district, similar to the work by Abrams et al. (2018), and that there may be varying behavior and propensities of the same judge in different districts, as seen by the work on judges in their home districts by Carr and McClain (2019).

For sentence types observed as binary indicators, we employ a linear probability model estimated using OLS to construct propensity indexes[9]. Fitted values of the LP model are estimated probabilities, and we calculate simple residuals as the difference between observed sentence type and the probability of that sentence type for each district-month combination. One exception to use of linear probability is for our measure of the worst type given, which is a categorical variable rather than a binary variable. We employ a multinomial logit for constructing this instrument. For measures of sentence intensity and motion approval (both continuous variables), we use OLS. For the residuals method, these first stage regressions are on district-case month dummies. We then regress the residuals on judge fixed effects using OLS. The first stage is given as:

$$M^n_i = T_i\alpha + D_i\beta + T_iD_i\lambda + \varepsilon_i, \tag{2.3}$$

where T_i is a vector of case month dummies and D_i is a vector of district dummies. We then predict $\widehat{M^n_i}$ to get the district-case month mean and calculate residuals for use in the judge fixed-effects regression:

$$M^n_i - \widehat{M^n_i} = Z_i\pi_n + \varepsilon_i \tag{2.4}$$

[9]We have also estimated district-month means using logit and probit models as a robustness check, but the effects were minimal. We preferred to use the linear probability OLS model due to it being the norm in the severity index literature.

and, using the procedure outline above, estimate our propensities, p_j^n as

$$p_j^n = Z_i'\widehat{\pi_n}. \tag{2.5}$$

For the comparison method, we run two sets of regressions of our measures, first with judge fixed effects and district-case month dummies and then with just district-case month dummies. We then use the procedure outline above to estimate our propensities, p_j^n, as

$$p_j^n = \widehat{M_{j,full}^n} - \widehat{M_{j,district-month}^n}. \tag{2.6}$$

In addition to constructing indexes using the two methods describe above for our eight measures, we also considered several robustness checks on our construction. The structured sentencing grid determines which type of sentences (active, intermediate probation, community probation) are possible for each given offense class-prior record point combinations. As a result, for each case we can observe whether specific sentence types were even possible. As a robustness check, we re-construct our index for active incarceration including a dummy for whether active incarceration was a possible sentencing type.

As discussed above, it is also of interest whether judges behave differently when sentencing to active incarceration as opposed to probation. To investigate this, we constructed our severity index for sentence intensity using the case-level worst maximum minimum, $p^{intensity,wmax}$, conditioning on whether the sentence was active incarceration or probation. We take two approaches to consider this. First, we use the residuals method to run regressions of $M^{intensity,wmax}$ on district-case month indicators and regressions of the fitted values on judge dummies for subsamples of cases by whether they were sentenced to active incarceration or probation. We then predict for the full sample on these subsample regressions to construct judge sentence intensity severity indexes by sentence type. Alternatively, we run the first-stage regression of $M^{intensity,wmax}$ as normal, but interact judge dummies with our measure of the worst sentence type given, $M^{type,worst}$ ("Robust" specification). In this regression, judge fixed effects can vary by sentence type. The difference between these two methods is practical. The former allows us to have judge severity indexes by sentence type (incarceration or probation) estimated for the same observation. The latter gives us a single severity index for each observation that represents the judges tendency to sentence given the type of sentence (incarceration or probation) in that case.

Finally, we consider two different types of leave-one-out estimators. Following the logic in Carr and McClain (2019), if prosecutors are able to match cases with judges the severity index is not exogenous to case characteristics that determine matching. One way to reduce the influence of prosecutorial discretion in matching cases to judges is to consider the judge's severity in all other districts as a proxy for severity in each district. We construct a leave-one-out estimator at the district-level, so that the estimated propensity of judge j in district d, $p_{j,d}^{n,loo}$ is the predicted p_j^n from all districts other than d. As is the norm, we will also estimate all of our core propensities using a leave-one-out-estimator so that there is no mechanical relationship between each observation's outcomes and the severity index[10]. Table 2.2 provides an overview of the eighteen severity indexes used in the main body of this work.

2.4.3 IV estimation

After constructing our measures and judicial severity indexes, we use these as instruments in a two-stage least squares estimation of the effect of incarceration and sentence length on recidivism. As discussed above, our measure of recidivism comes from matching the North Carolina DPS Offender Public Information data with our initial analysis data from ACIS. Aware of the limitations of this measure (in particular the limitations of using only those with active incarceration for recidivism), we estimate the effect of incarceration or sentence length (in 30-day month equivalents) on the likelihood of recidivism and on the total number of charges an offender commits, if they re-offend.

We are interested in estimating the effect of sentence type or sentence duration on the likelihood of recidivism:

$$Y = \beta_0 + \beta_1 X + \gamma \mathbf{\Gamma} + \delta + \psi + \omega + \varepsilon, \tag{2.7}$$

where Y is recidivism or the number of charges a re-offender commits (equal to zero if recidivism is not observed), X is either an indicator for sentence type (incarceration, probation) or a measure of sentence duration (total sentence, total incarceration spell), Γ is a vector of case and defendant controls, δ are crime-type fixed effects, ψ are district fixed effects, and ω are rotation period fixed effects. We are primarily

[10]At present we have not completed all leave-one-out estimations (with 145,242 observations in our full sample, leave-one-out estimators at the case-level take a long time to run). Preliminary results indicate the leave-one-out estimator has only moderate effects.

Table 2.2.
Descriptions of indexes used in main body of work

Index		Description
Incarceration, Residuals	$p^{active,res}$	Propensity to sentence to incarceration calculated using the residuals method
Incarceration, Comparison	$p^{active,comp}$	Propensity to sentence to incarceration calculated using the comparison method
Incarceration, Residuals Robust	$p^{active,res,robust}$	Propensity to sentence to incarceration calculated using the residuals method, controlling for whether an active sentence was possible at case level.
Incarceration, Residuals district LOO	$p^{active,res,loo}$	Propensity to sentence to incarceration calculated using the residuals method, estimated leaving out each district
Intermediate probation, Residuals	$p^{inter,res}$	Propensity to sentence to intermediate probation calculated using the residuals method
Intermediate probation, Residuals, multinomial logit	$p^{inter,res,ml}$	Propensity to sentence to intermediate probation calculated using the residuals method and a multinomial logit on the measure $M^{type,worst}$
Community probation, Residuals	$p^{inter,res}$	Propensity to sentence to community probation calculated using the residuals method
Community probation, Residuals, multinomial logit	$p^{inter,res,ml}$	Propensity to sentence to community probation calculated using the residuals method and a multinomial logit on the measure $M^{type,worst}$
Intensity relative to case worst max, Residuals	$p^{wmax,res}$	Propensity to sentence to a higher intensity relative to case worst max minimum possible using residuals method
Intensity relative to case worst max, Comparison	$p^{wmax,res}$	Propensity to sentence to a higher intensity relative to case worst max minimum possible using comparison method
Intensity relative to case worst max, Residuals conditioned on incarceration	$p^{wmax,res,inc}$	Propensity to sentence to a higher intensity relative to case worst max minimum possible using residuals method, conditioned on active incarceration
Intensity relative to case worst max, Residuals conditioned on probation	$p^{wmax,res,prob}$	Propensity to sentence to a higher intensity relative to case worst max minimum possible using residuals method, conditioned on any type of probation
Intensity relative to case worst max, Residuals robust	$p^{wmax,res,robust}$	Propensity to sentence to a higher intensity relative to case worst max minimum possible using residuals method controlling for worst sentence type
Intensity relative to case worst max, Residuals District leave-one-out	$p^{wmax,res,loo}$	Propensity to sentence to a higher intensity relative to case worst max minimum possible using residuals method and estimated leaving out each district level
Intensity relative to case best max, Residuals	$p^{bmax,res}$	Propensity to sentence to a higher intensity relative to case best max minimum possible using residuals method
Intensity relative to charge worst max, Residuals	$p^{max,res}$	Propensity to sentence to a higher intensity relative to charge worst max minimum possible using residuals method
Intensity relative to case worst min, Residuals	$p^{wmin,res}$	Propensity to sentence to a higher intensity relative to case worst min minimum possible using residuals method
Motions, Residuals	$p^{motions,res}$	Propensity to approve a motion for continuation using the residuals method

interested in the value of the coefficient β, which indicates the effect of sentence type or duration on the likelihood of re-offending, a major question in the literature on deterrence, incapacitation and optimal sentencing.

At issue is that the likelihood of incarceration, sentence duration and recidivism are likely correlated with unobserved (to the econometrician) defendant characteristics. In addition, the construction of our recidivism measure introduces selection bias due to the fact that defendants initially sentenced to probation or incarceration are more likely to be incarcerated if they re-offend, and therefore more likely to be identified as having re-offended in our data. As a consequence, the error term in Equation 2.7 is correlated with X and our estimate of β, $\hat{\beta}$, is biased. To obtain an unbiased estimate of β, we use two-stage least squares with a linear probability model. In the first stage we use our measures of judicial severity as instruments for the endogenous observation of sentence type or duration:

$$X = \alpha_0 + \alpha_1 p^n + \varphi\Gamma + \Delta + \Psi + \Omega + u. \tag{2.8}$$

We then use the predicted values $\hat{X}$ in place of the endogenous variable X:

$$Y = \beta_0 + \beta_1 \hat{X} + \gamma\mathbf{\Gamma} + \delta + \psi + \omega + \varepsilon. \tag{2.9}$$

For the two-stage least squares result to yield unbiased estimates of β, it is necessary that judicial propensity to sentence be correlated with sentence type and duration (relevance) and that judicial severity not be correlated itself with the error term ε (validity). The former requirement can be tested, and results are given below. The latter requirement, i.e. that the instrument can be excluded from the main regression, is more difficult to test.

As discussed in Carr and McClain (2019), it is possible that prosecutors use knowledge about the rotation of judges in North Carolina to strategically schedule cases. If this is the case, judge assignment is not truly random. In this case, it is possible that prosecutors will match judges to cases based on both the judge's proclivities and defendant and case characteristics. For the same reason that X was endogenous above, then, so too would p^n. Even beyond this issue, if judges are not randomly assigned to cases, the initial estimation of judge severity indexes would be biased.

For the purposes of this paper, we assume that relevance is satisfied. As this paper is not primarily presenting results as causal, but is instead comparing results, we proceed as if the instruments are valid. However, we are interested in possible evidence of selection bias on judges. As a result, we pay close attention to the effect of basing our judicial severity index on observations of judge behavior in all other districts (our "district leave-one-out" estimator of p_j^n) and on using the instrument proposed in Carr and McClain (2019) for assigning judges to cases. As is discussed in greater detail below, the primary issue with both the district leave-one-out and the additional IV measures is one of precision. As a result, we hesitate to interpret insignificant results as clear evidence of selection. Nevertheless, the use of these alternative estimators of judicial severity can point to possible issues with using pure judge fixed effects without accounting for possible strategic scheduling by prosecutors.

2.4.4 Classifying judge types

As a final step to this work, we use the severity indexes we construct to create four, broad classifications of judge types. We use judge propensities to sentence to active incarceration and to community probation to classify judges into four, mutually exclusive types: (1) harsh, (2) rehabbers, (3) either-ors, or (4) balancers. In particular, we use the following rules to classify judges[11]:

1. **Harsh:** $p_j^{active,res} > 0$ and $p_j^{community,res} < 0$
2. **Rehabber:** $p_j^{active,res} < 0$ and $p_j^{community,res} > 0$
3. **Either-or:** $p_j^{active,res} > 0$ and $p_j^{community,res} > 0$
4. **Balancer:** $p_j^{active,res} < 0$ and $p_j^{community,res} < 0$.

A 'harsh' judge is one who is more likely than their peers to sentence a defendant to incarceration and less likely than their peers to sentence a defendant to community probation. A 'rehabber' is the exact opposite: less likely to incarcerate, more likely to give community probation. Cases heard by these two types combined represent more than three-fourths of our observations, with a plurality heard by rehabbers. An 'either-or' is a judge who is more likely than their peers to both incarcerate and sentence to community probation, perhaps suggesting someone who likes to strike a balance in use of corrections technologies. Finally, a 'balancer' is someone less

[11]In our sample there are no observations where p_j^n exactly equals zero, so all judges are classified using these rules without having to decide where judges exactly at zero should lie. We could also consider additional types who are close to zero, focusing our typology instead on judges who fall in the tails of the distribution, but aimed instead at simplicity in classification.

likely to sentence to either light or harsh penalties. In general, balancer judges are the most likely to sentence to intermediate probation and either-ors are the least likely, suggesting that balancers strike a balance of supervision and probation while either-ors will assign defendants to either low or total supervision.

We use this typology in two ways. First, we directly regress recidivism on judge type to assess the impact of assignment to different types on re-offending. Additionally, we estimate the probability of being assigned to these different types based on the week an offense was committed, seeing whether there exist any significant effects on probability of assignment based on judge type. In particular, we regress each judge type successively on two primary defendant characteristics, whether the defendant is black or whether the defendant is female, controlling for district-by-crime-type fixed effects and the week of a rotation that a defendant committed their offense. The latter follows from Carr and McClain (2019), and controls for selection into judge rotation windows[12].

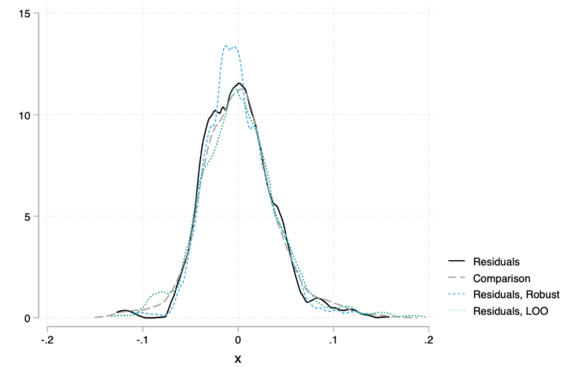
2.5 Results

2.5.1 Descriptive results

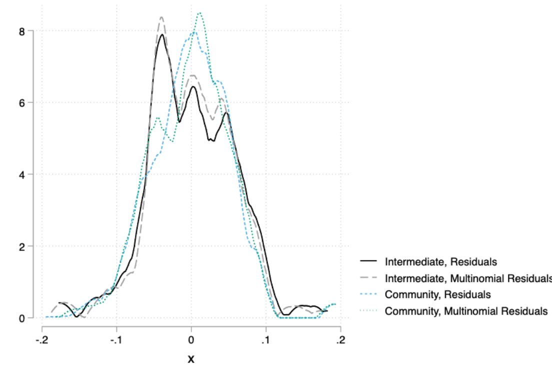
Figures 2.1(a), 2.1(b) and 2.1(c) give kernel density estimates of the distribution of main severity indexes for active incarceration, probation, and intensity of sentencing relative case worst max minimums, respectively. As is expected, the primary residuals method results are centered close to zero (representing mean leniency). Alternative specifications for active incarceration and probation have only modest effects, with the most noticeable effects being robust and leave-one-out estimates of propensity to sentence to active incarceration having fatter left tails. There is an interesting shift between sentencing to intermediate probation and community probation, with intermediate probation tending towards a trimodal structure with peaks at and on both sides of the zero center. As discussed above, this may represent the break-down of judge types around 'either-or' and 'balancer' judges.

For propensity to sentence relative to case worst maximum (Figure 2.1(c)), taking into account whether the sentence was incarceration or probation ("Robust" specification) shifts the distribution from unimodal to a multi-modal. The most interesting

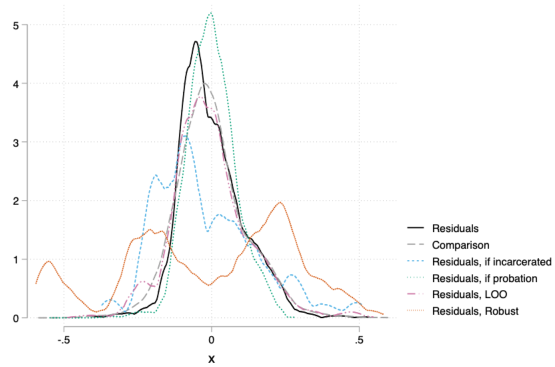
[12]This idea is part of on-going work with Dr. Jillian Carr, Dr. Mark Hoekstra from Texas A&M, and Dr. Daniel Berkowitz from the University of Pittsburgh.



(a) Four main incarceration severity indexes



(b) Four main probation severity indexes



(c) Six main sentence intensity (case worst maximum) severity indexes

Figure 2.1. Kernel-estimated densities of main severity indexes

feature of this distribution is the valley at zero. This result can also be seen when comparing the Robust specification to our other approach for taking into account propensities for probation and incarceration. When the sample is restricted to only cases resulting in probation, the distribution tightens significantly around zero. When the sample is then restricted to only cases resulting in active incarceration, mass shifts away from zero and the distribution widens significantly.

This result is expected for one mechanical reason and one potential behavioral reason. Mechanically, moving up the structured sentencing grids from least severe charges to most severe has two consequences: first, the bands of possible sentence lengths increase. Second, the availability of probation as a sentence type decreases. As a result, cases more likely to lead to probation are likely to have tighter sentencing bands and lower sentencing variation all together. That said, our construction of sentence intensity partially accounts for this by dividing actual sentence by worst possible.

A possible behavioral reason for this difference lies in how judges may perceive probation versus how they perceive incarceration. A tighter distribution around zero suggests that judges do not vary greatly in their propensities to sentence. This may suggest that, when sentencing to probation, judges in general are more likely to give similar sentences relative to the worst possible sentence they could have given with the full set of charges. When sentencing to incarceration, however, judges have a wider range of propensities: for example some may view incarceration as severe in itself and tend towards lighter sentences relative to worst possible while others view the bar for incarceration as an indicator that the severity of the crime merits significant time in detention. The distribution for severity using the limited sample of only those incarcerated again exhibits a valley around zero, suggesting that there may be a hardening of judge types once a judge has decided (or been forced to accept) incarceration. If this is the case, we may expect that judicial severity indexes will be more informative for cases of incarceration. Likewise, we may find that severity indexes built only around the binary choice of incarceration may miss important information about judge types.

2.5.2 Balance and monotonicity test results

Tables 2.3 to 2.6 present results from balance tests on the full set of instruments under consideration. Tables 2.7 to 2.9 present results from the weighted average monotonicity tests proposed by Dobbie et al. (2018) and Norris et al. (2019). The balance tests include the normal set of defendant characteristics, including race, gender and age, and a set of dummies for offense types. In addition, we include three case characteristics not normally included: whether the case was heard by a judge in their election district, whether the case was a jury trial, and whether it was a plea[13]. For the balance tests, we also compare results including a control for sentence length, since there may be mechanical correlations with severity indexes and case characteristics through sentence length.

[13]The alternative is for a non-plea trial to be heard by a judge only, which is at the discretion of the defendant to request a judge trial as opposed to a jury trial if they choose not to plea.

Table 2.3.
Balance tests on propensity to incarcerate

	Residuals		Comparison		Residuals, robust		Residuals, LOO	
Female def.	-0.001***	-0.001	-0.001***	-0.001*	-0.001**	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Black def.	0.000	0.000	0.000	0.000	0.000*	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Under 30 years old	-0.000***	-0.000	-0.001***	-0.000*	-0.001***	-0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Home judge	0.002	0.002	0.004	0.004	-0.001	-0.001	-0.007	-0.007
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
drugs	0.001	0.001	0.002*	0.001	-0.001	-0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
property	0.000	0.000	0.000	0.000	-0.001	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
violent	0.006***	0.004***	0.006***	0.004***	0.003*	0.001	0.002	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
sex crime	0.005***	0.002**	0.006***	0.003**	0.002**	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Jury trial	0.012***	0.008***	0.015***	0.010***	0.010***	0.007***	0.008***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Plea	0.003**	0.002*	0.003**	0.002*	0.002	0.001	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	145247	145247	145247	145247	145247	145247	145247	145247
Control for sentence length	N	Y	N	Y	N	Y	N	Y

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively. Standard errors clustered at the district level reported in parentheses. All specifications include district-level fixed effects.

Table 2.4.
Balance tests on propensity to sentence to intermediate probation

	Residuals		Residuals, Multinomial Logit	
Female def.	-0.003***	-0.002***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Black def.	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Under 30 years old	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Home judge	0.018**	0.017**	0.017**	0.016**
	(0.008)	(0.008)	(0.007)	(0.007)
drugs	0.007***	0.006***	0.007***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)
property	0.005***	0.005***	0.005***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)
violent	0.014***	0.012***	0.013***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)
sex crime	0.009***	0.005**	0.008***	0.005*
	(0.002)	(0.002)	(0.002)	(0.002)
Jury trial	0.007*	0.001	0.007*	0.001
	(0.004)	(0.004)	(0.003)	(0.003)
Plea	0.006***	0.005***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	145247	145247	145247	145247
Control for sentence length	N	Y	N	Y

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively. Standard errors clustered at the district level reported in parentheses. All specifications include district-level fixed effects.

Table 2.5.
Balance tests on propensity to sentence to community probation

	Residuals		Residuals, Multinomial Logit	
Female def.	0.002***	0.002***	0.003***	0.002***
	(0.001)	(0.000)	(0.001)	(0.001)
Black def.	-0.001*	-0.000	-0.001**	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)
Under 30 years old	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Home judge	-0.007	-0.007	-0.008	-0.007
	(0.008)	(0.008)	(0.008)	(0.008)
drugs	-0.003*	-0.002	-0.003*	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)
property	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
violent	-0.010***	-0.007***	-0.010***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Jury trial	-0.011***	-0.004*	-0.010***	-0.004
	(0.003)	(0.002)	(0.003)	(0.003)
Plea	-0.000	0.001	-0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	145247	145247	145247	145247
Control for sentence length	N	Y	N	Y

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively. Standard errors clustered at the district level reported in parentheses. All specifications include district-level fixed effects.

Table 2.6.
Balance tests on propensity to sentence relative to worst max possible

	Residuals		Comparison		Residuals, if incarceration		Residuals, if probation		Residuals, robust	
Female def.	-0.004***	-0.002*	-0.005***	-0.002**	-0.004*	-0.000	-0.004***	-0.002***	-0.084***	-0.064***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.006)	(0.005)
Black def.	0.001	0.001	0.001	0.001	0.003*	0.002	0.000	-0.000	0.027***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
Home judge	0.026	0.025	0.031	0.029	0.040	0.037	0.019	0.018	0.064***	0.050***
	(0.016)	(0.016)	(0.019)	(0.018)	(0.029)	(0.029)	(0.012)	(0.012)	(0.018)	(0.015)
drugs	0.009***	0.008**	0.012***	0.010***	0.013*	0.011	0.008***	0.007***	0.055***	0.040***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.007)	(0.007)	(0.002)	(0.002)	(0.013)	(0.012)
property	0.003	0.003	0.005*	0.005*	0.001	0.001	0.005***	0.005***	0.068***	0.068***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)	(0.005)	(0.002)	(0.002)	(0.015)	(0.013)
violent	0.023***	0.017***	0.028***	0.021***	0.035***	0.025***	0.017***	0.013***	0.201***	0.142***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.007)	(0.007)	(0.003)	(0.003)	(0.019)	(0.017)
sex crime	0.013***	0.004	0.017***	0.007*	0.016***	0.002	0.011***	0.005**	0.155***	0.067***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)	(0.002)	(0.002)	(0.016)	(0.014)
Jury trial	0.039***	0.025***	0.047***	0.030***	0.076***	0.053***	0.017***	0.008*	0.211***	0.071***
	(0.006)	(0.005)	(0.006)	(0.005)	(0.011)	(0.009)	(0.005)	(0.004)	(0.021)	(0.021)
Plea	0.008***	0.005**	0.009***	0.005**	0.008**	0.003	0.005***	0.003**	0.317***	0.287***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(0.019)	(0.017)
Observations	145247	145247	145247	145247	145247	145247	145247	145247	145247	145247
Control for sentence length	N	Y	N	Y	N	Y	N	Y	N	Y

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively. Standard errors clustered at the district level reported in parentheses. All specifications include district-level fixed effects.

There are several notable issues with the balance tests of these indexes. In Table 2.3, most of the specifications pass balance tests on the primary set of defendant characteristics. That said, several fail on female defendants and defendants under 30 years old, although these coefficients are very small and are often addressed by controlling for sentence length. In addition, the specifications that do not control for the possibility of incarceration ("Residuals, robust") or district assignment ("Residuals, LOO") all fail on multiple types of offenses. Most striking, perhaps, is the consistent positive relationship between severe judges and jury trials. Across all specifications, including those controlling for sentence length (which is likely to be heavily correlated with sentence length), there is a positive and significant—statistically and economically–effect of severity on the likelihood of a case being heard by a jury trial. Even in the two best preforming indexes, $p^{active,robust}$ and $p^{actve,LOO}$, there is still a positive relationship between a case being heard by a jury and the judge's severity.

This result is mirrored in other balance tests. It is less pronounced when considering the propensity to sentence in intermediate probation (Table 2.4), but remains present. On the other hand, judges who are more likely to sentence to community probation (largely including judge types we eventually call 'rehabbers') are significantly less likely to preside over jury trials. The result is likewise preserved when considering severity by sentence intensity. Cases disposed by jury trials are more likely to be heard by judges that give higher sentences relative to the case worst possible, even when controlling for sentence length. The strongest effect comes from judges who tend to give higher sentences when sentencing to active incarceration.

The indexes most likely to fail the balance tests are those on sentence intensity. This suggests there are consistent and systematic differences in allocation of defendants and cases across these types of judges. It is possible that this could be improved with a leave-one-out estimator[14]. Since most severity indexes are built around the propensity to incarcerate, it is interesting that severity built around sentencing behavior suggests a more complicated picture of sorting and judge types[15].

[14]This is currently in process for the full set of indexes, but initial results on a limited set indicate no noticeable effect.

[15]It is possible this is an artifact of constructing sentence intensity, since structured sentencing involves relatively lumpy bands. That said, within felony sentencing bands judges have a significant amount of leeway, often deciding between more than a hundred alternative sentencing arrangements, not taking into account mitigation and aggravation. In addition, there are still balance issues when controlling for the grid cell of structured sentencing, which should control for how location on the grid impacts the measure. Another potential test for this issue would be exploiting the variation in time of structured sentencing grids to extract information about how judges respond to marginal shifts in the availability of sentence ranges.

Table 2.7.
Weighted average monotonicity test: Sentence type indexes on incarceration

	Full Sample	Case characteristics						
		Black Def.	Female Def.	Home Judge	Drugs	Property	Violent	Sex crime
A. Active incarceration								
Propensity to incarcerate, residuals	0.927***	0.925***	0.488***	1.083***	0.983***	0.941***	1.299***	1.276***
	(0.099)	(0.114)	(0.071)	(0.139)	(0.140)	(0.112)	(0.143)	(0.179)
Propensity to incarcerate, comparison	0.842***	0.827***	0.455***	1.007***	0.883***	0.861***	1.161***	1.168***
	(0.078)	(0.084)	(0.060)	(0.108)	(0.111)	(0.091)	(0.125)	(0.181)
Propensity to incarcerate, residuals robust	0.980***	1.006***	0.533***	1.195***	1.020***	1.013***	1.351***	1.330***
	(0.104)	(0.119)	(0.076)	(0.145)	(0.146)	(0.118)	(0.154)	(0.195)
Propensity to incarcerate, residuals leave-one-out	0.230***	0.240***	0.080	0.307***	0.218***	0.224**	0.416***	0.671***
	(0.061)	(0.063)	(0.054)	(0.104)	(0.071)	(0.084)	(0.137)	(0.197)
B. Intermediate Probation								
Propensity to probation (inter.), residuals	0.127**	0.100	0.030	-0.014	0.126*	0.213***	0.335***	0.251*
	(0.053)	(0.066)	(0.048)	(0.101)	(0.066)	(0.063)	(0.088)	(0.142)
Propensity to probation (inter.), comparison	0.119**	0.097	0.027	-0.041	0.122*	0.198***	0.311***	0.215*
	(0.046)	(0.059)	(0.042)	(0.104)	(0.061)	(0.060)	(0.071)	(0.128)
Propensity to probation (inter.), residuals robust	0.144**	0.118*	0.043	-0.005	0.148**	0.231***	0.373***	0.288*
	(0.056)	(0.069)	(0.050)	(0.119)	(0.071)	(0.064)	(0.093)	(0.150)
Propensity to probation (inter.), residuals ML	0.123**	0.090	0.026	-0.043	0.121	0.216***	0.336***	0.284*
	(0.058)	(0.073)	(0.053)	(0.106)	(0.074)	(0.068)	(0.096)	(0.148)
B. Community Probation								
Propensity to probation (comm.), residuals	-0.443***	-0.444***	-0.240***	-0.535***	-0.456***	-0.483***	-0.704***	-0.558***
	(0.062)	(0.067)	(0.055)	(0.134)	(0.074)	(0.081)	(0.097)	(0.136)
Propensity to probation (comm.), comparison	-0.397***	-0.403***	-0.199***	-0.525***	-0.398***	-0.447***	-0.640***	-0.493***
	(0.052)	(0.057)	(0.049)	(0.125)	(0.064)	(0.070)	(0.081)	(0.128)
Propensity to probation (comm.), residuals robust	-0.441***	-0.438***	-0.232***	-0.515***	-0.449***	-0.489***	-0.689***	-0.570***
	(0.063)	(0.068)	(0.055)	(0.138)	(0.074)	(0.083)	(0.105)	(0.146)
Propensity to probation (comm.), residuals ML	-0.435***	-0.437***	-0.231***	-0.515***	-0.452***	-0.477***	-0.691***	-0.597***
	(0.062)	(0.067)	(0.055)	(0.127)	(0.075)	(0.079)	(0.091)	(0.134)

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively. Standard errors clustered at the district level reported in parentheses. All specifications include case controls and district-level fixed effects.

Table 2.8.
Weighted average monotonicity test: Sentence type indexes on probation only

	Full Sample	Case characteristics						
		Black Def.	Female Def.	Home Judge	Drugs	Property	Violent	Sex crime
A. Active incarceration								
Propensity to incarcerate, residuals	-0.555***	-0.625***	-0.013	-0.636**	-0.809***	-0.617***	-0.916***	-0.824***
	(0.131)	(0.120)	(0.128)	(0.247)	(0.153)	(0.179)	(0.115)	(0.184)
Propensity to incarcerate, comparison	-0.542***	-0.585***	-0.055	-0.655***	-0.758***	-0.615***	-0.823***	-0.820***
	(0.100)	(0.090)	(0.100)	(0.193)	(0.121)	(0.133)	(0.106)	(0.172)
Propensity to incarcerate, residuals robust	-0.680***	-0.774***	-0.135	-0.838***	-0.832***	-0.726***	-0.979***	-0.964***
	(0.127)	(0.131)	(0.136)	(0.244)	(0.152)	(0.180)	(0.114)	(0.193)
Propensity to incarcerate, residuals leave-one-out	-0.117	-0.176**	0.084	-0.003	-0.171**	-0.085	-0.354***	-0.656***
	(0.097)	(0.075)	(0.127)	(0.168)	(0.078)	(0.139)	(0.086)	(0.184)
B. Intermediate Probation								
Propensity to probation (inter.), residuals	0.203***	0.189**	0.330***	0.353**	-0.049	0.046	-0.014	0.289*
	(0.072)	(0.076)	(0.067)	(0.139)	(0.069)	(0.085)	(0.086)	(0.151)
Propensity to probation (inter.), comparison	0.156***	0.144**	0.268***	0.312**	-0.052	0.002	-0.028	0.170
	(0.056)	(0.065)	(0.045)	(0.135)	(0.065)	(0.064)	(0.069)	(0.148)
Propensity to probation (inter.), residuals robust	0.188**	0.173**	0.325***	0.342**	-0.065	0.030	-0.025	0.285*
	(0.072)	(0.077)	(0.070)	(0.154)	(0.074)	(0.086)	(0.094)	(0.164)
Propensity to probation (inter.), residuals ML	0.232***	0.221**	0.359***	0.406***	-0.035	0.063	0.005	0.290*
	(0.078)	(0.083)	(0.073)	(0.142)	(0.074)	(0.093)	(0.092)	(0.156)
B. Community Probation								
Propensity to probation (comm.), residuals	0.306***	0.314***	0.045	0.496***	0.456***	0.426***	0.472***	0.185
	(0.065)	(0.077)	(0.070)	(0.126)	(0.068)	(0.075)	(0.099)	(0.149)
Propensity to probation (comm.), comparison	0.274***	0.296***	0.020	0.510***	0.398***	0.400***	0.418***	0.174
	(0.054)	(0.066)	(0.059)	(0.110)	(0.061)	(0.062)	(0.087)	(0.142)
Propensity to probation (comm.), residuals robust	0.298***	0.298***	0.037	0.472***	0.457***	0.436***	0.464***	0.203
	(0.067)	(0.077)	(0.073)	(0.143)	(0.068)	(0.077)	(0.104)	(0.159)
Propensity to probation (comm.), residuals ML	0.286***	0.298***	0.025	0.451***	0.452***	0.407***	0.449***	0.201
	(0.065)	(0.077)	(0.069)	(0.125)	(0.070)	(0.073)	(0.098)	(0.147)

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively. Standard errors clustered at the district level reported in parentheses. All specifications include case controls and district-level fixed effects.

Table 2.9.
Weighted average monotonicity test: Sentence intensity indexes on sentence duration

	Full Sample	Case characteristics						
		Black Def.	Female Def.	Home Judge	Drugs	Property	Violent	Sex crime
Sentence to case worst possible max, residuals	15.704***	16.493***	12.626***	14.910***	19.016***	9.204***	28.698***	12.473***
	(1.779)	(2.418)	(1.455)	(2.666)	(2.333)	(1.127)	(5.033)	(3.424)
Sentence to case worst possible max, comparison	14.223***	14.780***	11.310***	13.771***	17.254***	8.510***	25.931***	11.174***
	(1.544)	(2.091)	(1.213)	(2.602)	(1.942)	(0.999)	(4.701)	(3.110)
Sentence to case worst possible max, if incarceration	6.656***	7.432***	5.069***	6.779***	8.700***	3.291***	12.819***	5.751***
	(1.023)	(1.385)	(0.879)	(1.724)	(1.367)	(0.642)	(2.734)	(1.983)
Sentence to case worst possible max, if probation	19.827***	20.072***	16.108***	16.746***	22.161***	13.223***	35.588***	15.572***
	(2.444)	(3.444)	(1.819)	(2.892)	(3.122)	(1.457)	(7.384)	(4.801)
Sentence to case worst possible max, robust	16.813***	17.162***	14.542***	16.738***	19.361***	10.838***	29.824***	23.133***
	(0.777)	(0.711)	(0.816)	(1.481)	(1.033)	(0.447)	(1.998)	(1.766)

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively. Standard errors clustered at the district level reported in parentheses. All specifications include case controls and district-level fixed effects.

In general, the core indexes under consideration in this work pass the weighted average monotonicity test of Norris et al. (2019). Table 2.7 give the results for the set of sentence type indexes on active incarceration. When using the district leave-one-out estimator for propensity to incarcerate, female defendants may exhibit non-monotonic effects. While the sign remains positive, the coefficient is no longer significant. This confirms results from Carr and McClain (2019) that there may be potential issues with judge assignment by gender. The same issue can be seen in Table 2.8, but it appears to be not at issue when using indexes for sentence intensity. Nevertheless, the first stage effects do appear to be generally weaker for female defendants across the board, suggesting that there may be some issues with judge behavior by gender of defendant. A similar result is seen below in our results from regressing judge type on defendant race and gender.

While there appear to be systematic issues with some of our instruments, many appear well-behaved, especially across the commonly included characteristics in balance tests. As a result, judge severity indexes may appear to balanced and well-behaved, but there may be systematic divergences in the types of cases or defendants judges have across alternative severity measures. Papers that report only one index may overstate the lack of possible selection or monotonicity problems. The strongest indexes are those that include additional information on case assignment, sentencing type, or districts where a case is disposed. This again highlights the potential multidimensionality of judge severity and the importance of considering more than just district-by-case month fixed effects when estimating a judge's propensity to sentence.

2.5.3 IV estimation results

Figures 2.2 to 2.7 present local polynomial regressions of each index on a range of potentially endogenous variable. The corresponding OLS coefficients from the first stage of our two-stage least-squares estimation are given in Tables 2.10 to 2.13. The tables present the first-stage coefficients without any additional controls and with the full set of controls included in our two-stage least-squares estimation. Figures 2.8 to 2.12 plot the regression coefficients with standard errors clustered at the district level for the full range of IV estimates. Plotted coefficients are on the endogenous variable,

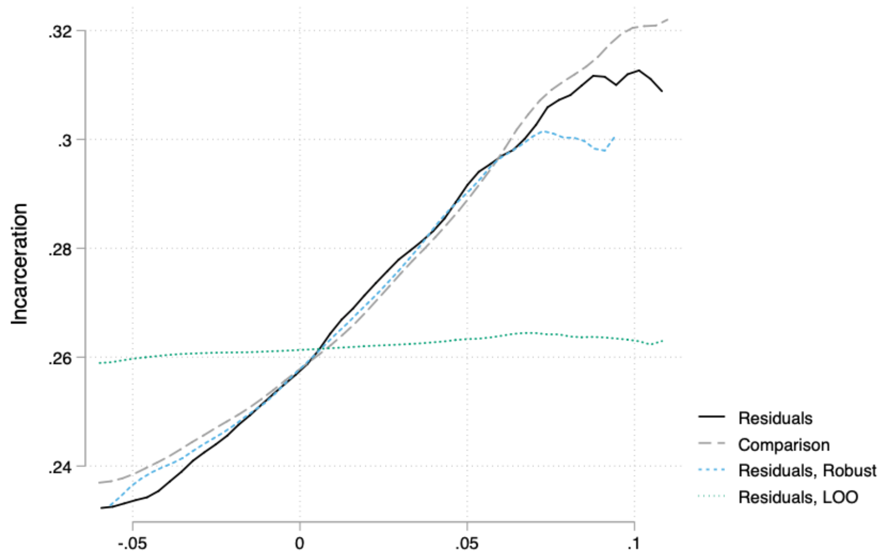


Figure 2.2. Propensity to incarcerate and likelihood of incarceration

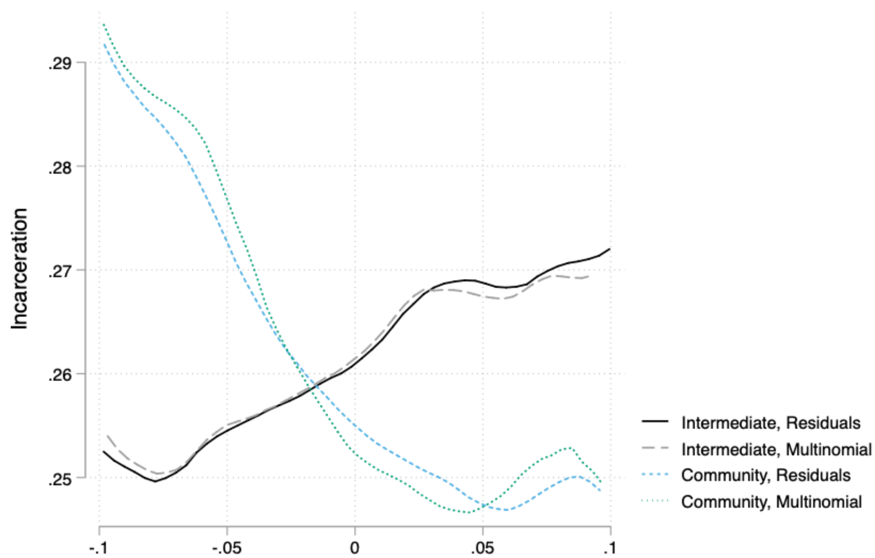


Figure 2.3. Propensity to sentence to probation and likelihood of incarceration

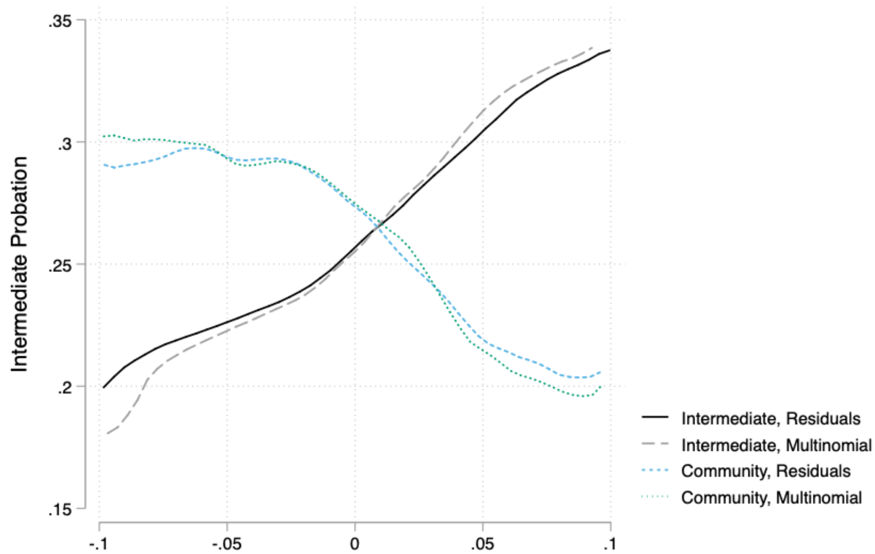


Figure 2.4. Propensity to sentence to intermediate probation and likelihood of intermediate probation

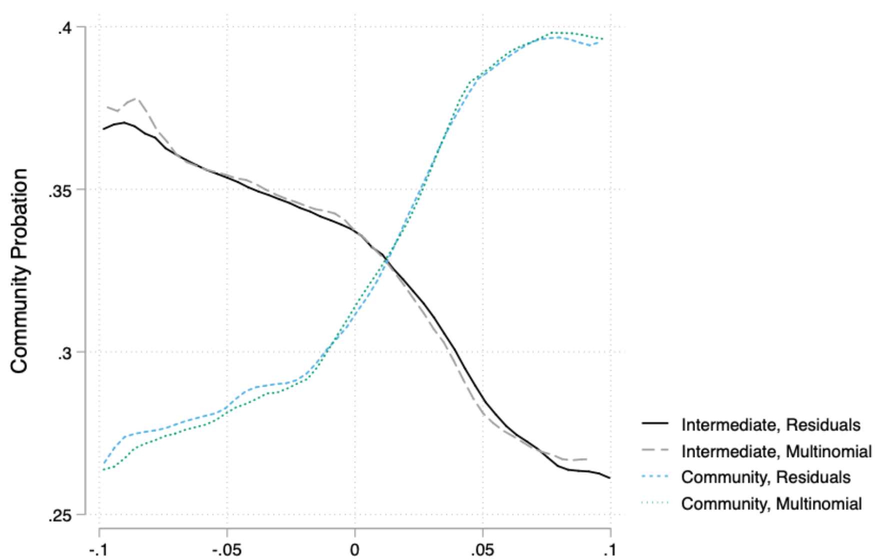


Figure 2.5. Propensity to sentence to community probation and likelihood of community probation

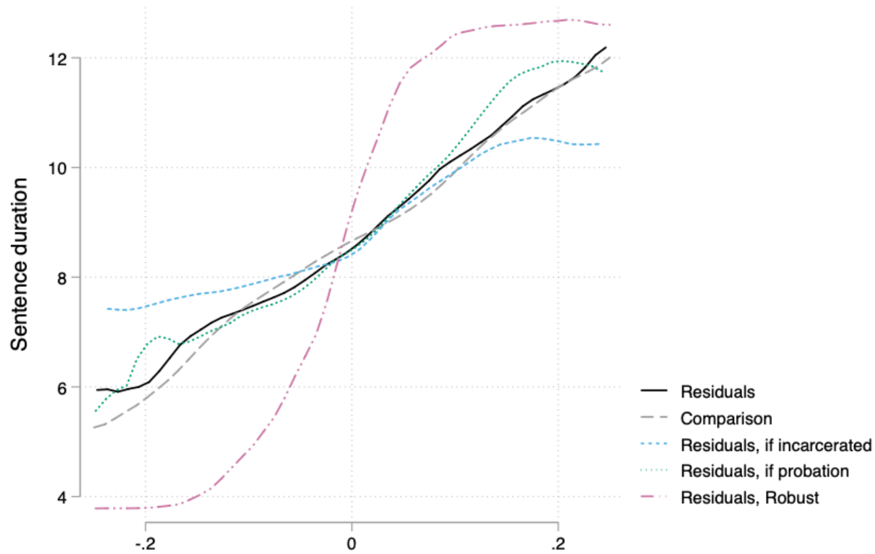


Figure 2.6. Propensity to sentence intensity and sentence duration (30-day month equivalents)

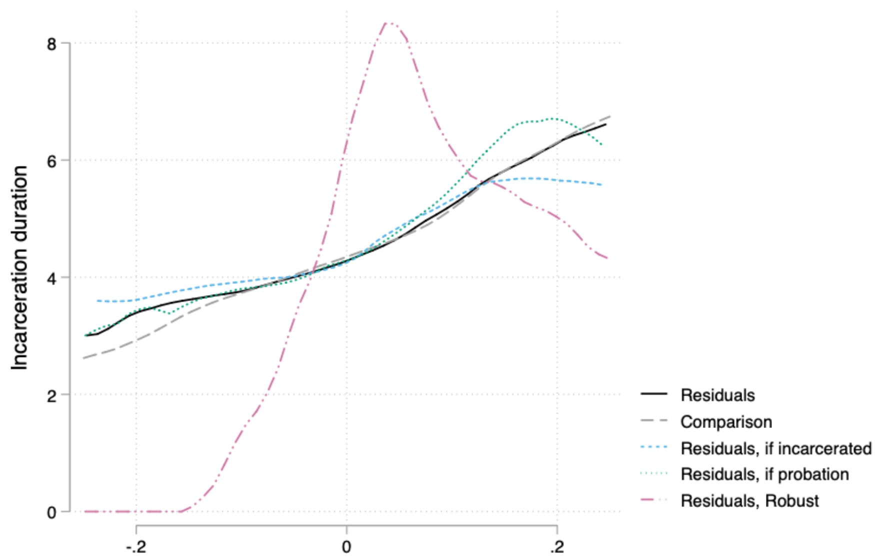


Figure 2.7. Propensity to sentence intensity and incarceration duration (30-day month equivalents)

Table 2.10.
First-stage coefficients of propensity to incarcerate instruments on actual incarceration

	Residuals		Comparison		Residuals, robust		Residuals, LOO	
Severity Index	0.946***	0.981***	1.000***	0.879***	0.992***	1.054***	0.124	0.255***
	(0.101)	(0.094)	(0.086)	(0.077)	(0.112)	(0.103)	(0.100)	(0.067)
Controls	N	Y	N	Y	N	Y	N	Y

Notes: Controls include all controls included in two-stage least squares (case and defendant characteristics), as well as district fixed effects. All standard errors clustered at the district level.

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Table 2.11.
First-stage coefficients of propensity to sentence to probation instruments on actual probation

	Intermediate, Residuals		Intermediate, Multinomial,		Community, Residuals		Community, Multinomial	
Severity Index	0.298***	0.325***	0.335***	0.363***	0.329***	0.294***	0.306***	0.276***
	(0.091)	(0.089)	(0.100)	(0.097)	(0.082)	(0.077)	(0.082)	(0.078)
Controls	N	Y	N	Y	N	Y	N	Y

Notes: Controls include all controls included in two-stage least squares (case and defendant characteristics), as well as district fixed effects. All standard errors clustered at the district level.

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Table 2.12.
First-stage coefficients of propensity to sentence intensity instruments on actual sentence duration (30-day months)

	Residuals		Comparison		Residuals, if incarcerated		Residuals, if probation		Residuals, robust	
Severity Index	16.450***	16.101***	16.979***	14.557***	7.190***	6.762***	20.111***	20.290***	17.305***	16.118***
	(2.098)	(1.773)	(1.708)	(1.540)	(1.145)	(1.030)	(2.920)	(2.424)	(0.758)	(0.709)
Controls	N	Y	N	Y	N	Y	N	Y	N	Y

Notes: Controls include all controls included in two-stage least squares (case and defendant characteristics), as well as district fixed effects. All standard errors clustered at the district level.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Table 2.13.
First-stage coefficients of propensity to sentence intensity instruments on actual incarceration duration (30-day months)

	Residuals		Comparison		Residuals, if incarcerated		Residuals, if probation		Residuals, robust	
Severity Index	11.467***	10.796***	11.794***	9.608***	5.401***	4.902***	12.477***	12.054***	12.990***	11.840***
	(1.562)	(1.374)	(1.419)	(1.214)	(0.874)	(0.800)	(2.059)	(1.769)	(0.778)	(0.722)
Controls	N	Y	N	Y	N	Y	N	Y	N	Y

Notes: Controls include all controls included in two-stage least squares (case and defendant characteristics), as well as district fixed effects. All standard errors clustered at the district level.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

and each coefficient is from a separate regression using a different severity index as the excluded instrument[16].

As can be seen in the local polynomial plots of our instruments against the potentially endogenous variables, the instruments are broadly relevant for the set of endogenous variables. The notable exception is the district leave-one-out index for propensity to incarcerate, which is essentially flat. This is confirmed in Table 2.10, where the simple OLS coefficient on this index is weak and insignificant. This suggests that estimating the severity of judges in a given district using their set of cases in all other districts can fail to predict sentencing outcomes in that district. This is of particular interest if there is strategic judge assignment that differs by prosecutor, who varies by district. In that case, we would expect that the behavior of a judge outside of a district might fail to adequately capture their "propensities" within a district. While this result is suggested by Carr and McClain (2019), it is also possible that the primary issue is one of power. Most judges have a majority of their cases in their home district, so using other districts will greatly reduce the amount of information available for estimation of their behavior[17].

Also of interest is the strong and negative relationship between propensity to sentence to community probation and actual incarceration. We use this opposite effect to justify basing our judge types on propensities to incarcerate and propensities to sentence to community probation. Judges do appear to largely fall on one side or the other, so that judges who are more likely to sentence to community probation are also less likely to sentence to incarceration, and vice versa. We are also interested in comparing these two opposite-effect instruments in the two-stage least squares results, which are largely confirmed between the two.

Results using the motions index are not presented in a table, but first stage results are weak, with higher values of the index generally leaning towards less severe judges (by the usual standards). This was particularly true for using the motions index to predict sentence type. The first-stage coefficient on the motions index on active incarceration is only -0.06, significant at the 5% level. On total sentence duration, the coefficient was only -4.35 (significant at the 5% level), compared to 15.62 from

[16]Results using the index constructed by using the measure of total continuations granted by a judge are not included in coefficient plots. In general, it had a consistent sign but much larger variance.

[17]One test for this would be to restrict predictions to all cases a judge makes outside their election district. This does slightly strengthen the first stage using the district leave-one-out index, but only slightly. This should highlight the real possibility judges behave differently in and outside their home district

$p^{wmax,res}$ and 36.98 on $p^{active,res}$. Consistent with these results, judge type is a weak predictor of propensity to approve continuations. While rehabbers are more likely and harsh judges are less likely, neither result is statistically significant. It is likely that continuations are granted largely on sound legal process reasons, which are dispersed relatively equally across judges[18].

First-stage results for sentence intensity instruments are also generally strong and relevant. As seen above, there are notable effects when taking into account judge's severity when sentencing to active incarceration versus to some form of probation. While most of the sentence intensity measures have a consistent upward slope across their support, the "Residuals, Robust" instrument has more of a binary reaction, with indexes values further away from zero much more likely to lead to lower or higher sentence lengths. A similar result is seen when considering incarceration duration, although in this case the robust specification leads to an unexpected negative slope as it moves further to the right of zero. It is possible that part of this is estimating a nonparametric function in a part of the support with limited information. As shown in Table 2.13, the linear relationship of the robust specification with sentence duration is strong, positive, and of a similar magnitude to the other indexes. The weakest first-stage results for sentence intensity are those based on how judges sentence to incarceration, and the strongest are based on how judges sentence to probation. As highlighted above, this demonstrates there may be meaningful differences in the sentencing behavior of judges when giving active incarceration as opposed to giving probation.

Figures 2.8 to 2.12 plot the coefficients of interest from our IV estimations, using the full set of instruments[19] Plotted at the top of each figure is the coefficient of the potentially endogenous variable from naive OLS.

The sign of all IV estimates confirm estimates from naive OLS, although the magnitude varies quite significantly. For the effect of any incarceration on likelihood of re-offending, the estimates range from 5% from naive OLS to over 20% using judge's severity based on sentencing intensity when giving probation, a more than

[18]While it fails to serve as a strong instrument, this fact about total continuations may make it a worthwhile variable to control for legal features of a case in regressions on criminal justice system outcomes where unobserved case type matters.

[19]The district leave-one-out instrument for propensity to incarcerate is excluded in Figures 2.9 and 2.10 only to improve readability. In both cases the point estimate in both specifications was more than twice the size of the coefficient on $p^{active,res}$, but the 95% confidence interval was very wide relative to other coefficients. The coefficients on the set of instruments measuring propensity to sentence to intermediate probation are excluded from Figure 2.8 for the same reason.

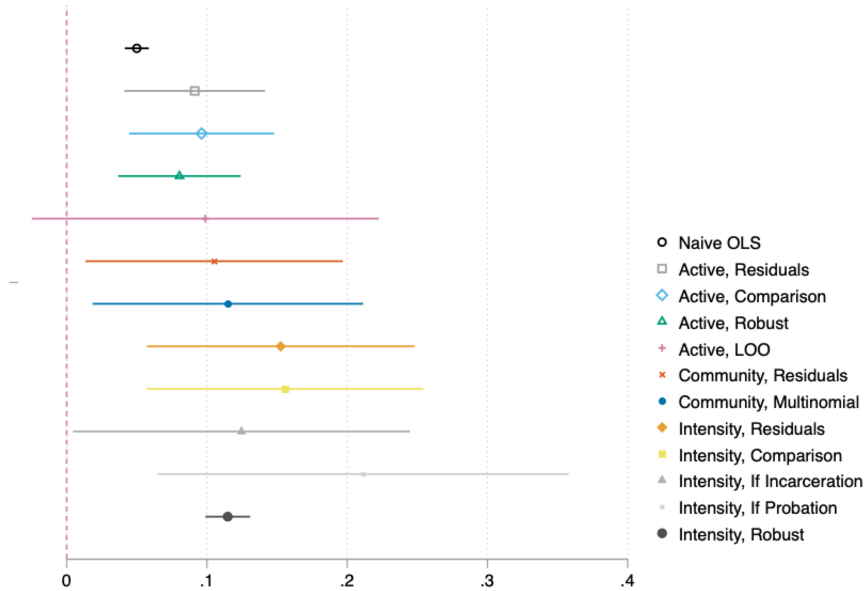


Figure 2.8. Two-stage least squares coefficients on instrumented variable across all instruments, effect of incarceration

300% increase. For the effect of an 30-day longer sentence, effects range from an 0.04% increase in the likelihood of re-offending under naive OLS to an over 0.4% increase, an over 900% shift in the effect. Even comparing only IV results, effects can range quite significantly. For active incarceration, effects range from around 8.5% to over 20%, a more than 150% difference. For community probation, IV coefficients range from just under -5% to just over 15%, a more than 200% swing.

While point estimates can vary, in general differences in estimates between instruments are not statistically significant. There are some exceptions, particularly when considering the effect of probation on recidivism and when using instrumenting using sentence intensity measures that distinguish between how judges sentence to incarceration and to probation. These differences again highlight the fact that there judge severity is likely multidimensional, and that there is value in distinguishing between how judges combine sentence types and length. In addition, even when taking into account the significant loss of power when instrumenting by judge severity, IV estimates are often statistically different from naive OLS estimates.

The vast majority of coefficients are statistically significant, the primary exception being the district leave-one-out estimator of propensity to incarcerate. As discussed above, since the point estimates of the LOO estimate is similar, the difference is likely

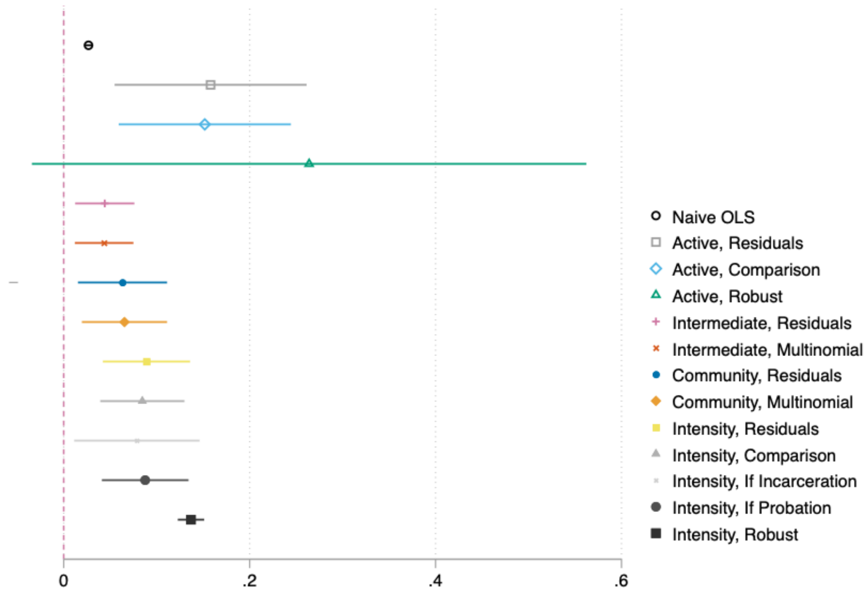


Figure 2.9. Two-stage least squares coefficients on instrumented variable across all instruments, effect of intermediate probation

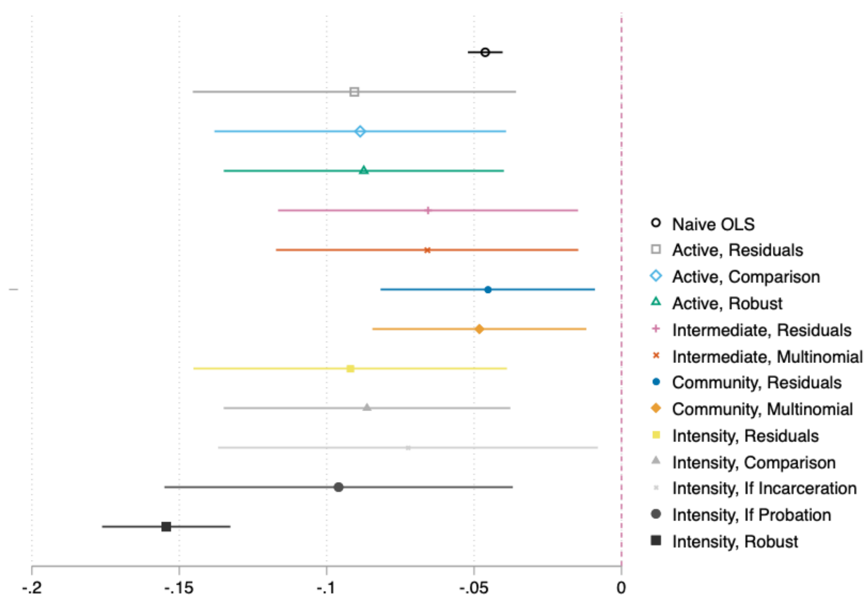


Figure 2.10. Two-stage least squares coefficients on instrumented variable across all instruments, effect of community probation

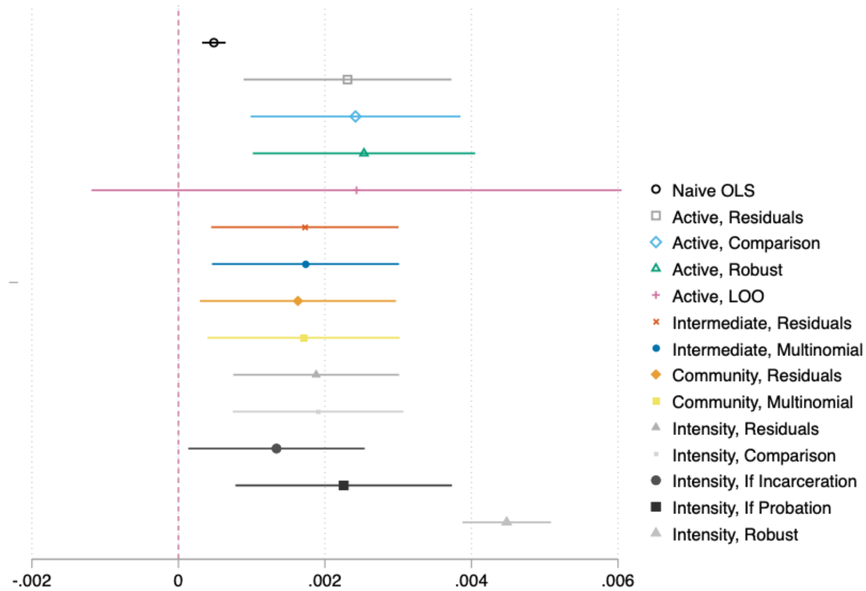


Figure 2.11. Two-stage least squares coefficients on instrumented variable across all instruments, effect of sentence duration

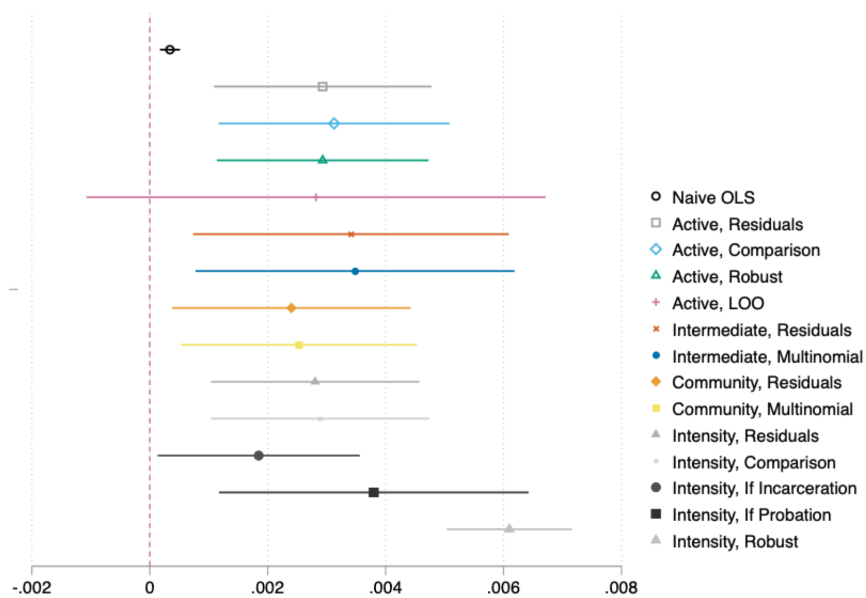


Figure 2.12. Two-stage least squares coefficients on instrumented variable across all instruments, effect of incarceration duration

largely driven by loss of power when estimating judge severity. At the same time, it again points to the possibility of bias in estimating judge severity due to selection of judges by prosecutors within a district.

While the impact of choosing a different severity index is not typically qualitatively different, the effects highlight the importance of treating point estimates from IV estimates using judicial severity indexes with caution. In general there is a wide range of possible effects, at times more than tripling the magnitude of effects by simply choosing a different severity index measure—measures that are broadly consistent and correlated with each other. More important than point estimates are confidence intervals. While the range of effects implied by these confidence intervals will also vary significantly across instruments, the degree of overlap can help mitigate overconfidence in estimates from a single instrument.

2.5.4 Judge type results

Tables 2.14 and 2.15 give results from our judge type analysis. In this analysis we limit our focus to four broad judge types—rehabbers, harsh, either-ors, and balancers. As defined above, we exploit the noticeable difference in behavior between judges likely to sentence to incarceration and to community probation to assign our typology. Across observations, 40.5% of cases are heard by rehabbers, 35.9% by harsh judges, 12.8% by balancers and 10.8% by either-ors. Across judges, 40.6% are harsh, 32.6% are rehabbers, 13.8% are balancers and 13% are either-ors.

Table 2.14.
Coefficients from regression of recidivism on judge types

	Rehabber		Harsh		Either-or		Balancer	
Judge type	-0.003*	-0.005**	0.004**	0.005***	-0.004*	0.003	0.002	-0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Observations	145247	145247	145247	145247	145247	145247	145247	145247
Controls	N	Y	N	Y	N	Y	N	Y

Notes: Controls include all controls included in two-stage least squares (case and defendant characteristics), as well as district fixed effects. All standard errors clustered at the district level.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Differences between shares of cases and shares of judges are likely explained mostly by regional differences in judge types and caseload: Democratic judges are 4.4% less likely to be harsh and black judges are 4.5% less likely to be harsh and 2.4% more

Table 2.15.
Likelihood of receiving judge type by defendant characteristic

	Rehabber		Harsh		Either-or		Balancer	
Black defendant	-0.008***		0.010***		0.001		-0.002	
	(0.003)		(0.003)		(0.002)		(0.002)	
Female defendant		0.015***		-0.014***		0.001		-0.002
		(0.005)		(0.005)		(0.002)		(0.003)
Observations	145247	145247	145247	145247	145247	145247	145247	145247

Notes: All regressions include district-by-crime-type fixed effects and offense commit date week of rotation fixed effects. All standard errors clustered at the district level.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

likely to be rehabbers when controlling for district fixed effects and judge gender. Both of these judge characteristics are more likely to be located in urban population centers in North Carolina, where a larger number of cases will originate. Female judges, who are more geographically dispersed, are 1.8% more likely to be harsh and 1.8% less likely to be rehabbers.

Table 2.14 present coefficients for the effect of judge type on recidivism. Cases heard by rehabbers are less likely to result in re-offense, while cases heard by harsh judges are more likely, even when controlling for a full set of case and defendant characteristics. Cases heard by either-or judges have less defendants who recidivate, but the effect is reversed and statistically insignificant when controlling for case and defendant characteristics. While the effects are small, the average rate of recidivism is only 8.5%, meaning the two main effects represent roughly 5.9% of the average each.

Table 2.15 presents the likelihood of receiving each judge type by two main defendant characteristics: race and gender. In this specification, we regress each judge type on defendant race or gender while controlling for district-by-crime-type fixed effects and the week of a rotation that a defendant *committed* their crime. We control for the week of a rotation period when a defendant committed their crime as an exogenous shifter to the set of judge schedules available for a defendant. Under the assumption that defendants are not choosing to commit a crime based on the likelihood of judge assignment, including these week-by-week effects captures the range of possible judge assignments (Carr and McClain, 2019).

The effects are striking, even if the magnitudes are small. Black defendants are 1% more likely to receive a harsh judge and .8% less likely to receive a rehabbers, even though these results are unlikely given the regional variation in harsh and rehabber judges and black defendants. Female defendants face an exactly opposite condition: they are 1.4% less likely to have their case heard by a harsh judge, but 1.5% more likely to have their case heard by a rehabber.

These results, combined with the discussion above of female defendants in balance tests, highlight the possible selection of judges by defendant characteristics. Even when controlling for the set of possible rotation windows, district make-ups, and the possible "specialization" by crime type of judges, we observe systematic differences in assignment of judge type to observable defendant characteristics. Notably, we find no systematic assignment across either-ors or balancers, both judge types that are likely to be less predictable to prosecutors. These results highlight the importance of pushing beyond simple balance tests, where race was often not identified as a potential issue, and point to a shadow of a doubt on "random" assignment.

2.6 Conclusion

In this chapter we compare a wide range of different judicial severity indexes used as instrumental variables in a two-stage least squares framework studying the effects of incarceration, probation, and sentence duration on recidivism. While we find broadly consistent effects with a significant overlap in their confidence intervals, we highlight that point estimates from a single instrument are likely to overstate their confidence. Combining the full range of possible effects from a full set of possible severity measures indicates that the range of possible effects can be quite large, at times leading to differences in point estimates of over 300%. Considering the full range of confidence intervals can lead to an extremely wide range of possible effects.

Most notable are differences in effects when controlling for judge behavior on multiple dimensions, for example considering how judges sentence to incarceration versus probation or considering the differences in judges who prefer incarceration and those who prefer community probation. Judge behavior, including judge severity, is very likely to be multidimensional, and a single instrument that does not include any information on these important dimensions may fail to be monotonic or balanced. In addition, they may lead to biased estimates of effects.

Beyond that, simple balance tests may fail to detect selection issues. By considering a richer set of instruments, we were able to develop a typology of judge type by severity that points to possible, systematic differences in judge assignment by race and gender and, consequently, systematic differences in the likelihood of recidivism. Works that only include information about propensity to incarcerate may miss these differences, and as a consequence may miss potential biases in their coefficient estimates. We strongly encourage future researchers to take advantage of the decreasing cost of computational power to include a broader range of propensity indexes when employing the methodology to estimate effects of random assignment of a third-party scorer or judge.

2.7 Appendix

Table 2.16.
Summary statistics on measures used in propensity index construction

	mean	sd	min	max
Type, active incarceration	0.260	0.439	0.000	1.000
Type, intermediate probation	0.277	0.448	0.000	1.000
Type, community probation	0.353	0.478	0.000	1.000
Sentence intensity, midpoint to case worst max	0.490	1.198	0.000	168.000
Sentence intensity, midpoint to case best max	2.508	10.195	0.000	607.500
Sentence intensity, midpoint to sentenced charges worst max	0.610	1.034	0.000	47.000
Sentence intensity, minimum to sentenced charges worst max	0.648	1.064	0.000	42.000
Sentence intensity, maximum to sentenced charges worst max	0.781	1.115	0.000	52.000
Sentence intensity, midpoint to sentenced charges best max	0.628	1.071	0.000	47.000
Sentence intensity, midpoint to case worst min	1.998	39.653	0.000	7560.000
Sentence intensity, midpoint to case best min	62.034	229.285	0.000	7560.000
Sentence intensity, midpoint to sentenced charges worst min	4.987	20.608	0.000	705.000
Sentence intensity, mininum to sentenced charges worst min	5.718	22.190	0.001	630.000
Sentence intensity, maximum to sentenced charges worst min	5.973	22.216	0.001	780.000
Sentence intensity, midpoint to sentenced charges best min	5.478	21.995	0.000	1290.000
Motions, total continuations	0.910	1.637	0.000	32.000

Table 2.17.
Summary statistics on sentence type indicators

	mean	sd	min	max
Propensity to incarcerate, residuals	0.001	0.036	-0.258	0.158
Propensity to incarcerate, comparison	0.001	0.040	-0.202	0.229
Propensity to incarcerate, residuals robust	0.001	0.035	-0.247	0.144
Propensity to incarcerate, residuals leave-one-out	0.001	0.043	-0.324	0.198
Propensity to probation (inter.), residuals	-0.000	0.059	-0.290	0.182
Propensity to probation (inter.), comparison	0.000	0.066	-0.393	0.377
Propensity to probation (inter.), residuals ML	-0.000	0.055	-0.269	0.174
Propensity to probation (inter.), residuals robust	-0.001	0.055	-0.265	0.207
Propensity to probation (comm.), residuals	-0.000	0.054	-0.290	0.555
Propensity to probation (comm.), comparison	-0.000	0.061	-0.294	0.331
Propensity to probation (comm.), residuals ML	-0.000	0.054	-0.280	0.564
Propensity to probation (comm.), residuals robust	-0.000	0.053	-0.223	0.478
Propensity to approve an additional continuation motion	-0.001	0.112	-1.172	1.577

Table 2.18.
Summary statistics on sentence intensity (max) indicators

	mean	sd	min	max
Sentence to case worst possible max, residuals	-0.001	0.100	-0.538	0.701
Sentence to case worst possible max, comparison	-0.001	0.111	-0.538	0.851
Sentence to case worst possible max, if incarceration	-0.011	0.196	-0.939	2.247
Sentence to case worst possible max, if probation	-0.000	0.072	-0.431	0.721
Sentence to case worst possible max, residuals leave-one-out	-0.005	0.117	-0.538	0.959
Sentence to case best possible max, residuals	0.007	0.664	-1.784	7.505
Sentence to charge worst possible max, residuals	0.000	0.071	-0.563	0.285
Sentence to charge worst possible max, comparison	0.001	0.081	-0.563	0.499
Minimum sentence to charge worst possible max, residuals	-0.000	0.068	-0.437	0.391
Maximum sentence to charge worst possible max, residuals	-0.000	0.078	-0.555	0.386
Sentence relative to charge best possible max, residuals	0.000	0.074	-0.566	0.319

Table 2.19.
Summary statistics on sentence intensity (min) indicators

	mean	sd	min	max
Sentence to case worst possible min, residuals	-0.010	1.304	-5.060	13.453
Sentence to case worst possible min, comparison	-0.010	1.454	-7.341	15.578
Sentence to case worst possible min, if incarceration	-0.133	4.214	-18.549	41.525
Sentence to case worst possible min, if probation	-0.008	0.747	-2.181	15.191
Sentence to case best possible min, residuals	0.112	14.961	-59.830	96.889
Sentence to charge worst possible min, residuals	-0.010	1.349	-4.307	7.300
Sentence to charge worst possible min, comparison	-0.014	1.478	-8.001	9.029
Minimum sentence to charge worst possible min, residuals	0.007	1.660	-5.316	9.624
Maximum sentence to charge worst possible min, residuals	0.007	1.632	-5.200	9.740
Sentence relative to charge best possible min, residuals	-0.010	1.518	-4.861	7.401

Table 2.20.
Weighted average monotonicity test: Sentence length indexes on incarceration

	Full Sample	Case characteristics						
		Black Def.	Female Def.	Home Judge	Drugs	Property	Violent	Sex crime
Sentence to case worst possible min, residuals	0.005***	0.006**	0.002	0.003	0.006**	0.004	0.012***	-0.000
	(0.002)	(0.002)	(0.001)	(0.004)	(0.002)	(0.003)	(0.003)	(0.007)
Sentence to case worst possible min, comparison	0.005***	0.005**	0.001	0.003	0.006**	0.005**	0.010***	0.001
	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)	(0.003)	(0.006)
Sentence to case worst possible min, residuals conditioned on incarceration	0.001**	0.001	0.000	0.001	0.002**	0.001	0.003**	-0.000
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Sentence to case worst possible min, residuals conditioned on probation	-0.000	0.001	0.003	-0.008	-0.001	-0.001	0.001	-0.000
	(0.004)	(0.005)	(0.003)	(0.005)	(0.004)	(0.005)	(0.008)	(0.009)
Sentence to case best possible min, residuals	0.001***	0.001***	0.001***	0.001**	0.002***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Sentence to charge worst possible min, residuals	-0.006***	-0.005**	-0.003	-0.004	-0.006**	-0.007***	-0.015***	-0.010*
	(0.002)	(0.002)	(0.002)	(0.005)	(0.003)	(0.002)	(0.003)	(0.005)
Minimum sentence to charge worst posible min, residuals	-0.006***	-0.005***	-0.003*	-0.004	-0.006**	-0.007***	-0.014***	-0.010**
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)	(0.004)
Maximum sentence to charge worst possible min, residuals	-0.006***	-0.005**	-0.003*	-0.004	-0.006**	-0.007***	-0.014***	-0.010**
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.005)
Sentence relative to charge best possible min, residuals	-0.005***	-0.004**	-0.002	-0.002	-0.005*	-0.005**	-0.012***	-0.008
	(0.002)	(0.002)	(0.002)	(0.005)	(0.003)	(0.002)	(0.003)	(0.005)

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively. Standard errors clustered at the district level reported in parentheses. All specifications include case controls and district-level fixed effects.

CHAPTER 3. THE IMPACT OF NEW FARMER ENTRY ON FARM CAPITAL AND FEDERAL PROGRAM PARTICIPATION: EVIDENCE FROM THE LAND CONTRACT GUARANTEE PROGRAM

3.1 Introduction

Efforts to incentivize beginning and new farmers entry into agriculture have been ongoing since at least the 1990s, with funds and the number of programs meant to support beginning and socially disadvantaged farmers increasing through the 2002, 2008, 2014, and the most recent 2018 farm bills. With roughly a quarter of current operators in the United States above the age of 65 and given the difficulties in securing capital and land necessary to launch a new farm, policymakers have been concerned with increasing rates of retirement and the entry dynamics of farm operators. At the same time, there is a variety of different types of new farmers, with differences age, previous experience, financial standing, and education all likely to play important roles in determining their risk preferences, financial viability, and technology adoption behavior. Similarly, new farmers are often touted as more likely to participate in local food systems and seek out innovative practices, technologies, and crops, although this effect likely varies significantly by the age and experience of new entrants. As new farmers are incentivized to entry and existing farmers retire, then, there are likely to be changing patterns of farm behavior, which may have important consequences on the use of agricultural support policies, the food system, and farm incomes. This work addresses the impact of new farmer entry on two primary outcomes: farm capital investment (total machinery assets, new tractors, total tractors) and the participation of new farmers in government-sponsored conservation programs.

Understanding the dynamics of entry, in particular the types of new farmers incentivized by different programs and their impact on local food systems, is an important component in evaluating the success of these programs and in predicting future changes to the farm economy. Estimating these dynamics, however, is made difficult by the endogeneity of entry and exit choice and the impact of the broader local economy on entry and production decisions. This research contributes to this understanding by considering the rich set of policies introduced between 2002 and 2014 to incentivize new and beginning farmers and combining multiple datasets, including the 2002, 2007, 2012, and 2017 farm censuses, data on county-level unemployment, and data on producer prices from the Bureau of Labor Statistics. Constructing a first stage estimate of how different policy types changed the distribution of farmer types by age and experience across counties allows for the impact of new farmer entry on capital investment, participation in conservation programs, use of fertilizer, and participation in crop insurance.

While previous studies have considered the factors determining financial success of new farmers or the household and operational characteristics of new farmers in snapshots of time, the financial constraints and characteristics of new farmers will be endogenous to the economic and market conditions that influence entry and exit decisions. This research uses variation in the timing of policy implementation across counties, as well as variation across the definition of who qualifies for different new and beginning farmer programs to identify shifts in the distribution of new and beginning farmers by county. In particular, we exploit variation in exposure to the Land Contract Guarantee Program through its pilot study, which offered a limited number of federal loan guaranties a year in six states (IN, IA, ND, OR, PA, and WI) from 2002 to 2012. Beginning in 2013, the Land Contract Guarantee Program was extended nationwide. To the extent that support for land purchases of new farmers have different effects on operators with different risk and technology adoption preferences, we expect to find significant effects from the implementation of a program on technology adoption decisions, participation in conservation programs, use of fertilizers on cropland, and the percentage of acres in a county that are insured under federal crop insurance programs.

We find a significant positive effects from land contract guarantees on the fraction of principal operators in a county that have less than five or ten years of experience. In addition, we consider entry and exit dynamics by comparing successive agricultural census waves to calculate exit of farmers with zero to nine years of experience and farmers with five plus years of experience. The nature of the land contract guarantee program is that retiring farmers are incentivized to sell their land to new, non-family farmers. While the program supports land sales between retiring and new, it does not on its own provide any support for the first crucial years of a new operator's business. As a result, we are interested in comparing the effect of exit rates for new farmers (farmers with less than five years experience) and existing farmers (farmers with greater than five years experience). We find that the land contract guarantee has a negative and significant effect on exit rates of established farmers, but a small and insignificant positive effect on the exit of farmers who were in their first five years of experience in the proceeding census year. In addition, we find mixed results on the effect of fraction of new farmers in a county on participation in conservation programs and capital investment.

Using first-stage results from our preferred estimation, we find that counties with higher fractions of new farmers have a lower percentage of cropland and pasture and a lower total acreage under federal conservation programs; lower total machinery assets and machinery assets per operation; more tractors over all, but a lower percentage of tractors that are new; lower fertilizer expense per acre of cropland, and a lower (but statistically insignificant) percentage of cropland under federal crop insurance programs. Under an alternative specification including the post-2012 nationwide roll-out of the Land Contract Guarantee Program we find some diverging results, particularly higher participation in conservation programs and larger total assets. We believe these results may be a consequence of the expansion of alternative federal incentives for new and beginning farmers in the 2014 Farm Bill, especially policies that supported the sale of conservation land to new farmers and policies aimed at improved capital investment on Conservation Reserve Program land. This research is of interest both to researchers studying farmer entry or farm management, and to policy-makers interested in downstream effects of efforts at encouraging new farmer entry. The methodology employed can be expanded to consider a wider range of federal and state incentives for farmer entry, exploiting regional and time variation to capture the effect of farmer entry on farm management decisions.

3.2 Background

3.2.1 Academic literature

The topic of new and beginning farmers has been of interest to both academics and policymakers for several years. Previous studies have considered the characteristics of beginning farmers and the dynamics of farm entry. Many have considered the household and operational characteristics of new farmers in snapshots of time, for example Ahearn and Newton (2009) and Bigelow et al. (2016). Some have focused instead on the determinants of new farmer success, for example Mishra et al. (2009); Ahearn (2011); Katchova and Ahearn (2016). These find capital constraints, the availability of land, and written business plans are the most important drivers of success. Multiple studies have highlighted the difference between operator age and entry into farming (Boehlje, 1973; Gale, 2003). These have found that new farmers are not always young farmers, and that many new principal operators have extensive experience in agriculture either as laborers or as the children of farmers.

Perhaps most similar is (Katchova and Ahearn, 2016), who use three agricultural census rounds to construct a longitudinal dataset in order to examine the dynamics of farmland ownership and leasing among young and beginning farmers. They find significant interactions between age and beginning farmers, with younger farmers more likely to expand farm size more rapidly. In general, roughly a quarter of principal farm operators in the United States are over the age of 65, with a national average of 59 years for sole operators (Ahearn and Newton, 2009). As a result, the dynamics of retirement and age are important. While "new" farmer is not synonymous with "young" (Boehlje, 1973), Katchova and Ahearn highlight important relationships between farm development and growth and the age of new farmers. They find that entrants into farming with higher ages tend to have smaller operations with less growth, with most of expansion in farm size and operations coming from young entrants. However, they also find that these younger entrants are most likely to be cash constrained, and recommend programs that expand access to loans for land purchase as opposed to rental or leases.

3.2.2 Policy

Congress and administrative agencies have sought to incentivize entry by easing financial, risk management, and knowledge constraints on new and beginning farmers through loan support through the Farm Service Agency (FSA) or the Farm Credit System (FCS), advantageous crop insurance terms through the Risk Management Agency (RMA), and programs such as the Beginning Farmer and Rancher Development Program (BFRDP), Section 2501 support for veterans and socially disadvantaged entrants, the Conservation Reserve Program Transitions Incentive Program (CRP-TIP). Importantly, the implementation of different incentives and incentive types vary in time (by the farm bill that funded them), space (through the use of pilot programs), and target (by differences between USDA and RMA definitions of new and beginning farmers and between special classes of new and beginning farmers).

To the extent that policies aimed at encouraging farmer entry or preventing failure by new farmers are successful, the different policy mixes should have had impacted the entry dynamics of American agriculture in a way that may have important consequences to the agricultural market and the food system in the next ten to twenty years, especially if the types of farmers encouraged to enter vary significantly from the types of farmers likely to exit over the time period of a policy intervention. This is true for comparing across policy types, for example comparing education policies aimed at inexperienced farmers and loan support for any new farmer. It is also true comparing policies within a type. For example, policies for new farmer education that focus on production practices or availability of other government support programs are likely to be more effective for new and beginning farmers with limited experience in agriculture as opposed to individuals who have worked on a family farm or as a farm laborer for multiple years. On the other hand, access to education on farm financial planning may be of use for all types of new farmers, even those who have multiple years of experience in production.

Land Contract Guarantee Program

In this paper we consider the Land Contract Guarantee Program (LCGP). In particular we focus on the impact of the pilot program of the LCGP, which allows for a classic difference-in-difference approach for estimating the entry of new farmers into a county. Beginning in 2002, six states had access to a limited number of loan guarantees from the USDA Farm Service Agency (FSA). In 2012, the LCGP was expanded nationwide following the success of the pilot program. We utilize agricultural census data from 1992 through 2017, which allows us to estimate the effect of the LCGP on new farmer entry in both the six initially treated states and nationwide following 2012.

The LCGP offers two primary guarantees to qualifying land contract sales. The first is a prompt payment guarantee (LCPP), which guarantees up to three annual installment payments, including the cost of any related real estate taxes and insurance. The LCPP ensures that the seller of agricultural land will be guaranteed an uninterrupted flow of payments, allowing up to three uses of the guarantee. The second is a standard guarantee (LCSG), which provides a 90% guarantee on the balance of a land contract. In the case of purchaser default, the seller can liquidate the real estate and receive 90% of the remaining principal. The seller may also choose to retake possession of the property under guarantee.

In order to qualify for a guarantee under the LCGP, the sale must meet several conditions. First, it must be a new land contract for a farm or ranch with a purchase price under $500,000 and the buyer must be able to meet a minimum down payment of 5% of the purchase price. Payments must be amortized over at least 20 years with equal installments, with a balloon payment allowed after ten years and a fixed interest rate for the first ten years. The interest rate is further restrained to not exceed the FSA direct farm ownership loan interest rate that is in effect when the guarantee is approved. Both the LCPP and the LCSG are in effect for at most ten years, which allows the guarantee to focus on the crucial initial period of farmer entry. The FSA does not charge a fee for participation in the guarantee program.

In order to qualify, the buyer must be a beginning or socially disadvantaged farmer. The Farm Service Agency considers a farmer to be beginning if they have less than ten years experience operator a farm or ranch and participate substantially in the operation of a farm or ranch now. For the purposes of loan support, which the LCGP represents, the new and beginning farmer must not already own a farm larger than 30% of the average acreage of farms in a county. We use county-level variation in 30% of the average farm size to investigate possible heterogeneous effects across farm types, assuming that larger averages allow for larger scale operations to qualify for loan guarantees. The USDA defines a socially disadvantaged farmer as a member of a group traditionally disadvantaged in agriculture, namely American Indians, Alaskan Natives, Asian Americans, African Americans, Native Hawaiians, other Pacific Islanders, Hispanics, and women. The buyer must also have a satisfactory credit history and be capable of obtaining credit absent the guarantee to qualify. Finally, the buyer must not be a family member of the seller.

The LCGP was first introduced as a pilot program in the 2002 Farm Bill. A pilot program was authorized for nine states, although the FSA eventually only rolled out a pilot in six: Indiana, Iowa, North Dakota, Oregon, Pennsylvania, and Wisconsin. The pilot program was launched from 2002 onwards and was subsequently made permanent and nationwide in the 2008 Farm Bill. While the program was made permanent in 2008, a phased roll-out meant that nationwide coverage was not achieved until 2012. The original pilot program only included beginning farmers, but socially disadvantaged farmers were added in 2008. From the initial 2002 pilot program until the present day, appropriations have been sufficient to fully fund all eligible proposals.

3.3 Empirical strategy and data

The primary interest of this research is the difference in farm management and participation in federal programs for farmers by the experience level of new farmers. The entry, behavior, and potential exit of new farmers are all likely to be closely related to each other and to additional dynamics in the farm economy. At the most basic level, the cropping and operational decisions of new farmers will have significant effects on their survival (Mishra et al., 2009). In addition, the entry of farmers will depend on the current state of the farm economy, the cost of land, the cost of capital, and opportunities for new farmers to market their products. As a result, estimating the relationship between farmer experience and farm operational decisions poses multiple problems of simultaneity and omitted variable bias, leading to likely endogeneity between farm operation variables and farm experience.

Unfortunately, we do not have farm-level data on the experience and operation decisions of farm operators with a rich set of covariates that allow us to construct a structural model of farm entry for estimation. Instead, we observe five-year snapshots of operator counts by experience through the United States Agricultural Census. In particular, we observe six waves of the agricultural census with county-level counts of farm operators by experience, with counts in bands of less than three years experience, three to four years experience, five to nine years experience, and greater than or equal to ten years experience. In addition to counts of operators, we use county-level agricultural census data on the total acreage of crop- and pasture-land in a county as control variables for the extent of agricultural activity in a county. We also observe several farm operation variables in the agricultural census, most notably the total value of machinery assets on farm operations in a county, the total count of tractors in a county, the total count of tractors less than five years old in a county, and the total expenses on fertilizer in a county. In addition, we observe the total acres under conservation programs and insured by federal crop insurance in a county. When appropriate, we transform variables to be per-acre or fraction of total acres in county (e.g. to obtain fraction of acres in county insured under federal crop insurance programs).

The USDA considers a farm to be a beginning farm if all operators on the farm have less than ten years of individual experience as an operator. As such, we can construct a dataset that has multiple measures of farm entry of new and beginning farmers at the county-level over time. First, we are able to observe the count of operators in a county who entered farming since the last census (farm experience less than 5 years). Second, we can observe the count of farm operations who qualify for USDA new and beginner farmer benefits (experience less than 10 years)[1] Finally, we observe an absorbing state of farm operators with ten or more years of experience. Of particular interest to us in this study is the fraction of total operations in a county with a principal operator with less than five years experience, since these represent farms that have entered since the previous census wave. In addition, we consider the share of farmers with less than ten years experience, since these are the set of farmers eligible for federal support through the LCGP.

Since we observe successive five-year waves of data on the same counties, we can also construct two net-exit variables: the number of farmers exiting in the first nine years of their experience and farmers exiting with five or more years experience. The reason we can only construct these two, overlapping variables is from the nature of observing five-year census waves. Consider comparing counts of farmers at difference experience levels in 1997 and 2002. Farmers with less than five years experience in 1997 who remain principal operators in 2002 will now have between 5 and 9 years experience during the 2002 census wave. If a farmer was observed in 1997 immediately after entering (0 years experience) but has exited by the time of the 2002 census, they will have exited in the first five years of experience. If a farmer was observed in 1997 just before having 5 years experience (4 years experience), and exist just before the 2002 census, they will have exited at just under 9 years experience. Farmers with five or more years of experience in 1997 who remain principal operators in 2002 will now have ten or more years of experience, placing them in the absorbing state. Any farmers who exited from the 5-9 or 10+ bands will have exited at five or more years of experience. To construct a net exit variable for census year y, we compare operator counts in year y, $O_{band,y}$, to the corresponding experience levels they would have had in $y-1$:

$$\text{Net Exit 0 to } 9_y = O_{0-4,y-1} - O_{5-9,y} \tag{3.1}$$

[1]In practice, the USDA requires that all operators on a farm have less than ten years experience on any farm. The agricultural census data we use observes the count of principal operators by years of experience on present operation. As a result, it should be seen as an upper bound on the count of qualifying new and beginner farmers. If a farm operation is classified as 'new and beginning' in our study, it may actually fail to qualify for USDA benefits. In our study, we are interested in whether participation in federal programs results in higher fractions of new farmers in a county. Some component of the share of new farmers in our data, then, would fail to qualify for USDA support. This should bias our results on the effect of introducing a new program towards zero.

and

$$\text{Net Exit 5 plus}_y = O_{5-9,y-1} + O_{10+,y-1} - O_{10+,y}. \tag{3.2}$$

3.3.1 First stage

As discussed above, the challenge with using observational data on the number of new farmers in a county to assess the impact of having more new farmers in a region on farm management and farm economy measures is that the entry, survival, and exit of farmers will depend on farm operation decisions and the local farm economy. To address this issue, we use variation in the timing of a policy to incentivize and support new farmer entry across six states in the United States to construct a difference-in-difference model for the fraction of new farmers in a region. This estimation will form the basis of a first-stage estimate of county-level fractions of new farmers out of total operators, which will then be used in a second-stage to estimate the effect of having a higher percentage of farmers in a county being new and beginning on several aggregate farm management decisions.

The basic model we are estimating is

$$W_{band,y,i} = \frac{O_{0-4,y,i}}{O_{total,y,i}} = \beta_0 + \beta_1 Post_y + \beta_2 T_i + \beta_3 Post_y \times T_i + \gamma \mathbf{\Gamma_{yi}} + \lambda \mathbf{\Lambda_y} + \varepsilon_{iy}, \tag{3.3}$$

where $Post_y$ is an indicator variable if the year is greater than 2002, T_i is an indicator variable if county i is in the pilot program for the LCGP that began in 2002, Γ_i is a vector of county-level time-varying controls and Λ is a vector of time-varying controls. We consider the 'post' period to be census waves following the year 2002, since at the time of the 2002 census the pilot program was not in effect even though it was authorized and funded that year. Γ contains controls for county-level area under crops and pasture, as well as county-level unemployment rates. Λ contains data from the Bureau of Labor Statistics on producer prices received for an aggregate of all farm products, as well as corn and soybean prices specifically[2]. This regression is also run using the number of operators with experience less than ten years ($O_{0-9,y,i}$) in the left-hand side numerator, since this captures the entire range of operators eligible for

[2] We also run our regressions with one-year lagged averages of producer price indexes on these products

the LCGP. Our primary focus is on the fraction of farmers with less than five years experience because this represents entry since the last census wave.

The core data we run this first-stage regression on is agricultural census data from 1992 to 2012, merged with county-level and time-varying data on economic and price conditions. We also have access to the 2017 agricultural census data, which follows the nation-wide roll-out of the Land Contract Guarantee Program in 2012. While we only have a limited amount of information on the effect of this roll-out, with only one census wave including the whole nation, we use an event-study approach with county-level and year fixed effects[3] to take advantage of this additional data:

$$W_{band,y,i} = \frac{O_{0-4,y,i}}{O_{total,y,i}} = \beta_0 + \nu_i + \xi_y + \sum_{k=1}^{5} \beta_k \mathbb{1}(K_{iy} = k) + \beta_6 \mathbb{1}(K_{iy} \geq 6) + \varepsilon_{iy}, \quad (3.4)$$

where ν_i are county-level fixed effects, ξ_y are year fixed effects, and K_{iy} is the number of periods prior to treatment (2007 for the pilot states, 2017 for the nation-wide states[4]). In effect the event-study approach traces the time-trend of each county relative to its assignment to treatment. In addition, we run a simple static model using only a pre-post comparison period:

$$W_{band,y,i} = \frac{O_{0-4,y,i}}{O_{total,y,i}} = \beta_0 + \nu_i + \xi_y + \beta D_{iy} + \varepsilon_{iy}. \quad (3.5)$$

Finally, we run a similar set of regressions using the net exit variables calculated above. To do so, we divide the next exit during years 0 to 9 and during years 5+ over the total number of operators in a county in year y. We also calculate a total net exit by adding up both net exit variables and likewise dividing it by the total number of operators. All fractions are multiplied by 100 to express in percent terms. We consider several robustness checks on our main first-stage regressions. First, we restrict our control groups to counties within the same agricultural statistical districts (ASDs) as the treatment counties. In addition, we restrict our control group even further to those states directly neighboring treatment states. For these checks we are hoping to preserve our main results, as we are refining comparison to a more directly

[3]Following Borusyak and Jaravel (2017) we also run the same design using random effects at the county-level. A Hausman test on the two sets of estimates and find no systematic difference between the two. We prefer to use the fixed effects model, since we believe it is likely that unit-level effects are correlated with independent covariates.

[4]We adjust the year of treatment as we did in the above difference-in-difference approach since the census data was collected in the year prior to treatment going into effect.

comparable control group. Finally, we construct treatment randomly using a random number generator at the state-level, then assign treatment to all states with a random number greater than the average random number.

3.3.2 Second stage

We then use the results of these first-stage estimates of fractions of new farmers to leverage policy variation to estimate credibly exogenous levels of new farmer activity in a local farm economy. Following estimation of Equations 3.3, 3.4 and 3.5, we estimate the relevant dependent variable (fraction of new farmers in a county). We then use this estimate as an independent variable in a series of second-stage regressions on the aggregate farm operation outcomes we are interested in:

$$Y_{iy} = \alpha_0 + \alpha_1 \widehat{W_{band,y,i}} + \xi \mathbf{\Gamma_{yi}} + \psi \mathbf{\Lambda_y} + \nu_i + \varepsilon_{iy}, \tag{3.6}$$

where Y_{iy} is a different outcome including fraction of acres in a county under federal conservation programs, total acres under federal conservation programs in a county, total machinery assets (million dollars) in a county, machinery assets (million dollars) per operator in a county, total tractors in a county, total tractors less than five years old in a county, fertilizer expenditures per acre of cropland in a county, fraction of acres in a county under federal crop insurance programs, and rent per acre; and ν_i are county-level fixed effects. In this regression we include controls for farm product, corn, and soybean prices on a 1-year lag.

3.4 Results

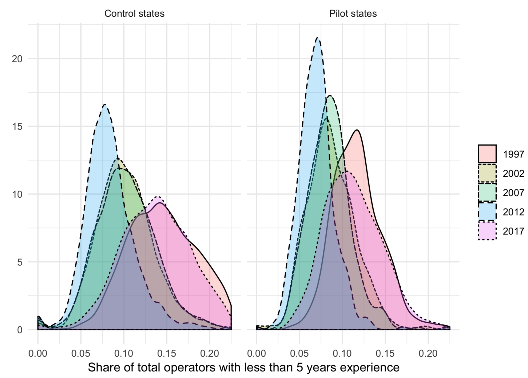
Descriptive statistics of county-level variables of interest are given in Table 3.1. The national average of fraction of new farmers in a county is 11.89%, with control states having a higher percentage overall. Table 3.7 in the Appendix gives the same set of summary statistics but with only neighboring states in the control group, showing more similar characteristics. Figures 3.1(a) and 3.1(b) show the kernel-estimated density of shares of total operators from 1997 to 2017 in control and pilot states. Despite differences in the means between these two groups, they have very similar densities over time, although pilot states see a slight abatement in the post-2002 decline relative to control states. This effect is captured by our difference-in-difference estimates given below, and constitutes the main results from our primary regression.

Table 3.1.
Summary statistics for control and pilot states

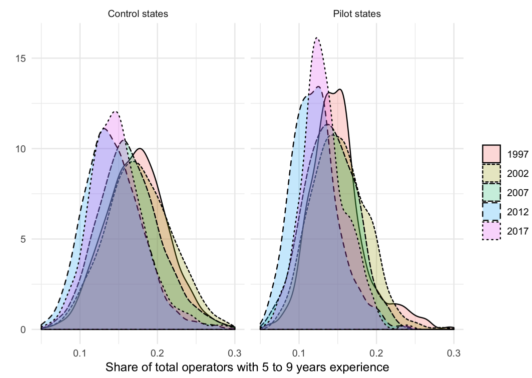
	Control	Pilot	Total
Total operators	694.20	883.78	720.04
% of operators < 5 yrs. exp.	12.19%	9.98%	11.89%
% of operators < 10 yrs. exp.	28.44%	23.94%	27.82%
Total acres cropland	83,033.83	138,876.48	90,644.61
Total acres pastureland	109,521.56	48,794.81	101,245.13
% acres under conservation programs	4.01%	4.55%	4.10%
Total acres under conservation programs	11,393.56	14,610.09	11,891.76
Total machinery assets (Million $'s)	57.48	98.35	63.06
Machinery assets per operator (Million $'s)	84,393.99	111,570.53	88,104.40
Total tractors in county	1,301.73	2,093.87	1,409.93
% tractors new	12.56%	8.79%	12.05%
Fertilizer expenses per acre ($)	42.99	45.73	43.37
% acres under federal crop insurance	48.60%	51.01%	48.95%
Rent per acre ($'s)	16.28	33.83	18.70

The decline in share of new farmers from 1992 to 2012 and rise in 2017 can also be seen in Figure 3.2, which shows the development of share of operators with less than five years or five to nine years experience over the entire data horizon. As can be seen, during the full data horizon (1992-2017), the share of new farmers saw a secular decline from 1997 to 2012, before picking up in the 2017 census. The share of operators with five to nine years experience increased to 2002, before also declining out to 2012 and raising again following the 2012 census. In between 2012 and 2017 saw a faster rise in the share of new farmers, which we believe reflects the increased intensity of incentives offered following the 2014 Farm Bill.

A similar but slightly more nuanced dynamic can be seen in Figure 3.3, which plots average net exit figures calculated in Equations 3.1 and 3.2 for control and pilot states over time. Solid lines are exits of farmers of experience levels of five or more years, dotted lines are exits of farmers with experience levels less than five years. Positive values represent counties that experienced more exits than entrants over that period, negative values represent a net entry. On average, there is net entry of farmers in the first nine years of experience and net exit of farmers with five or more years of experience. Similar to Figure 3.2, the net exit rates show that following 2012 there was a significant move towards net entry in both control and pilot states. While the



(a) Kernel-estimated densities of share of total operators with less than 5 years experience in control and pilot states



(b) Kernel-estimated densities of share of total operators with less than 5 years experience in control and pilot states

Figure 3.1. Share of total operators new densities

total fraction of new farmers is lower for pilot states, in general pilot states see a greater tendency towards net entry, although the difference is small and inconsistent.

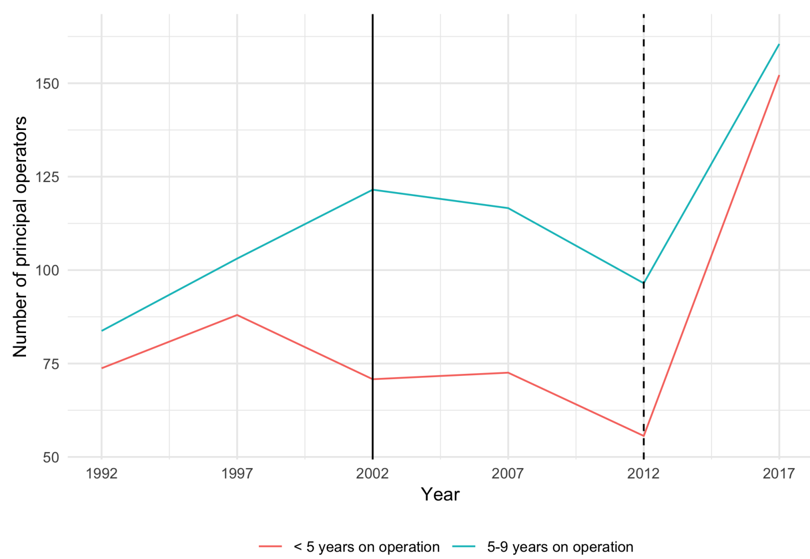


Figure 3.2. Mean fraction of farmers with <5 years and 5 to 9 years experience over time

Results on the effect of local farm economies on net exit can be seen in Figures 3.4 and 3.5. Figure 3.4 is a scatter plot of state-level unemployment averages against net exit rates, with a linear fit for control and pilot states in both the pre- and post-pilot periods. Pre-2007 shows a general negative trend between unemployment and net exit, although both are close to zero and insignificant. Post-2002, however, shows the more expected positive relationship between higher unemployment and higher net exit, with a steeper and significant effect for pilot states. Much of the variation is likely due to the fundamentally different economies from 1992-2002 and 2007-2017, and highlights the importance of controlling for local economic conditions.

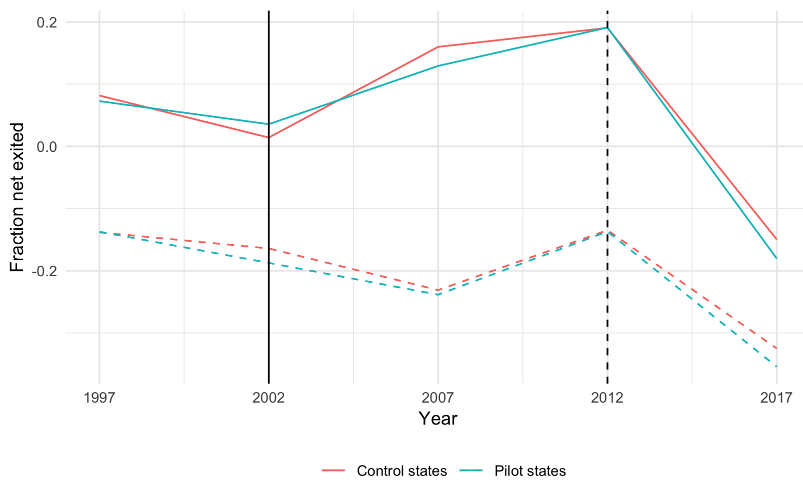


Figure 3.3. Net exit of farmers divided by total operators for control and pilot states over time.

Note: Dotted lines represent net exit of operators with experience of 0 to 9 years, solid lines represent next exit of operators with experience of 5 plus years.

Figure 3.5 plots state-level five-year average of land values against the same net exit rates, with a linear fit for control and pilot states in both the pre- and post-pilot periods. Relationships between the two are small and largely insignificant, but in both the pre- and post-pilot periods there is a positive relationship between the cost of land and exit of farmers. This highlights the importance of the LCGP in abating exit and facilitating entry, as the high cost of land is a major component of financial drag on new farm operations. Since few new farmers will be able to pay cash for land, higher land costs represent higher debt payments and more exposure to price risk given this higher leverage. Of note in this figure is the slight flattening of the relationship in pilot states in the post-pilot period, although this effect and difference is not statistically significant. In presentation of our second-stage results below we

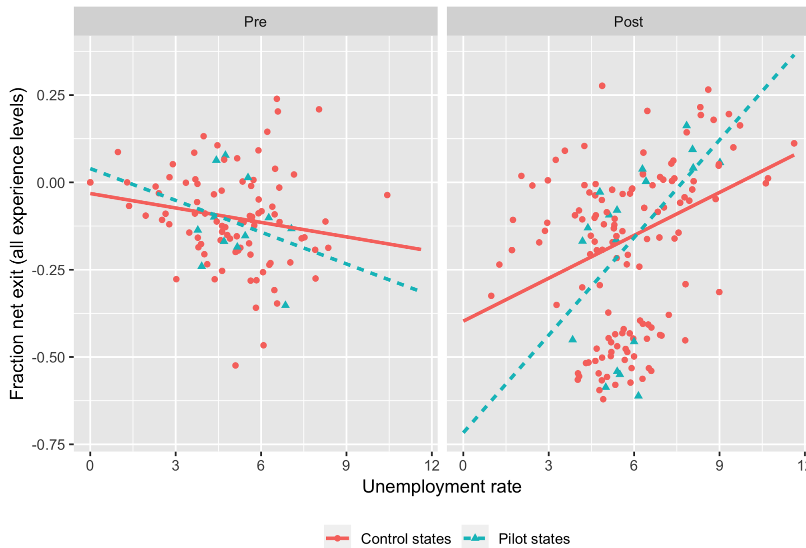


Figure 3.4. Scatterplot of state unemployment rate against net exit of farmers divided by total operators (all experience levels), control and pilot states in pre- and post-pilot program periods.

Figure 3.5. Scatterplot of five-year average land value in state against net exit of farmers divided by total operators (all experience levels), control and pilot states in pre- and post-pilot program periods.

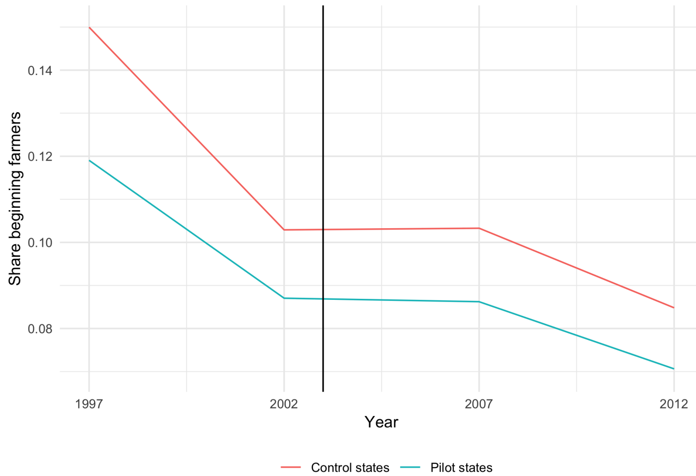
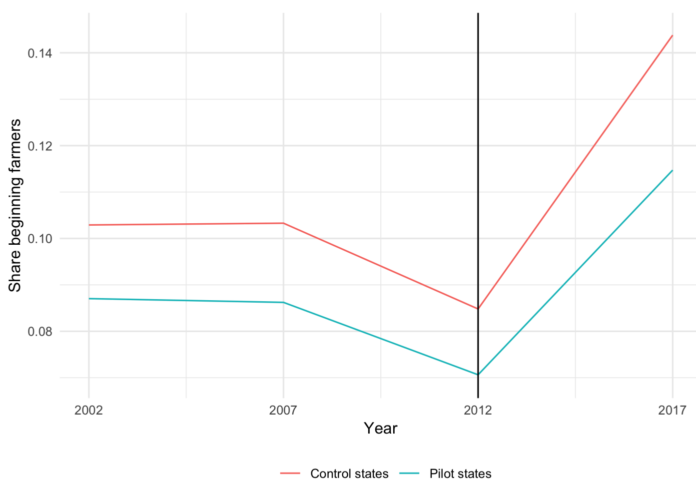


Figure 3.6. Parallel trends of share of total operators with less than 5 years experience in control and pilot states in main regression sample (1997-2012)



!h

Figure 3.7. Share of total operators with less than 5 years experience in control and pilot states following 2012 nation-wide roll-out (1997-2017)

address this issue from another angle, considering the impact of new farmer entry on rent-per-acre prices.

Our main regression considers a limited data horizon only extended out to 2012. As can be seen in Figure 3.6, control and pilot states demonstrate very similar development of new farmer shares across time, and appear to have largely parallel trends. While the post-2002 shift in policy in pilot states appear minor, the difference-in-difference results below suggest that pilot states experience an abatement of the secular decline in share of new operators out to 2012 equal to around 8.4% of the national average. Using results from the event study that includes the entire data horizon, the effect of treatment constitutes a roughly 69% abatement of the average trend downwards in 2007 and 2012. Figure 3.7 compares the trends of control and pilot states following the 2012 nationwide roll-out of the LCGP. As can be seen, the share of new farmers in control and treatment states expand following 2012, with the rise faster in control states. It is likely this reflects additional incentives in the 2014 Farm Bill, especially around conservation programs and crop insurance incentives which may have impact non-pilot states more.

3.4.1 First-stage results

Table 3.2 presents results for the main first-stage regression from Equation 3.3. Columns 1-3 present results including different subsets of county-varying and time-varying controls. Column 4 is our preferred specification, with county- and time-varying controls. For all first-stage regressions we calculate classic standard errors, Huber-White robust standard errors, standard errors clustered at the state-level, and, following Bertrand et al. (2004) and Rokicki et al. (2018), wild cluster bootstrapped standard errors (999 repetitions). For our main regression, there are no systematic divergences between these approaches for estimating confidence intervals and significance[5]. In our main regression tables we report standard errors clustered at the state level, as we find it likely that there may be non-independence in the error structure across states given the importance of state-level policymaking and implementation for assessing policy impacts.

Panel A in Table 3.2 presents results for the difference-in-difference on fraction of total principal operators that have less than five years experience. In other words, the fraction of total principal operators who are new since the previous census wave. As can be seen, there is a positive and significant effect on the main difference-in-difference coefficient (Post × Pilot) in all specifications, although significance falls

[5]We do observe differences in significance when regressing on the fraction of farmers with less than ten years experience, finding high statistical significance ($p<0.00$) for the difference-in-difference coefficient in all methods except for clustering at the state level. We still report the standard errors clustered at the state level in our tables.

Table 3.2.
First-stage difference-in-difference results for share of total operators with <5 years experience and with < 10 years experience.

	Share of total operators that are new			
	(1)	(2)	(3)	(4)
A. New is < 5 years experience				
Post 2002	-3.859***	-3.570***	-68.747***	-60.455***
	(0.240)	(0.283)	(4.306)	(7.547)
Pilot state	-2.379***	-2.001***	-2.381***	-2.033***
	(0.658)	(0.584)	(0.658)	(0.584)
Post × Pilot	0.790*	1.034**	0.792*	0.999**
	(0.448)	(0.480)	(0.448)	(0.484)
B. New is < 10 years experience				
Post 2002	-5.691***	-5.515***	-47.001***	-22.133
	(0.323)	(0.413)	(6.472)	(13.850)
Pilot state	-4.750***	-3.871***	-4.752***	-3.842***
	(1.429)	(1.206)	(1.429)	(1.196)
Post × Pilot	0.798	1.328	0.799	1.342
	(0.717)	(0.811)	(0.717)	(0.821)
Observations	15369	15369	15369	15369
County controls	N	Y	N	Y
Time controls	N	N	Y	Y

Notes: Standard errors clustered at the state level given in parentheses.

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

without including county-level controls (including unemployment and data on cropland and pasture, which reflect market conditions and supply of available land for the LCGP, respectively). In our main result, pilot states experienced a roughly 1% increase in the share of new farmers relative to control states. While this is a small effect, it represents roughly 8.4% of the national average and roughly 10% of the pilot-state average share of new operators. Panel B presents results for the share of total operators with less than ten years experience, which represents the qualifying group. The magnitude of the effect is larger, but significance using clustered standard errors falls just outside 10%[6].

Table 3.3 presents the result of three different robustness checks on our main difference-in-difference results for the share of total operators with less than 5 years experience. Panel A limits the sample to only those agricultural statistical districts

[6]The difference-in-difference effect on share of farmers with less than ten years experience is significant using classic standard errors, Huber-White robust standard errors, and bootstrapped standard errors.

Table 3.3.
Robustness checks on first-stage difference-in-difference results for share of total operators with <5 years experience.

	Share of total operators that are new			
	(1)	(2)	(3)	(4)
A. Sample limited to ASDs with treatment states				
Post 2002	-3.836***	-3.492***	-68.968***	-56.832***
	(0.229)	(0.257)	(4.536)	(7.475)
Pilot state	-2.268***	-1.854***	-2.270***	-1.892***
	(0.642)	(0.553)	(0.642)	(0.556)
Post × Pilot	0.768*	1.027**	0.770*	0.995**
	(0.442)	(0.478)	(0.442)	(0.481)
Observations	14503	14503	14503	14503
B. Control sample limited to neighboring states				
Post 2002	-3.353***	-3.081***	-67.143***	-68.060***
	(0.114)	(0.125)	(2.538)	(3.621)
Pilot state	-1.077***	-0.831***	-1.077***	-0.821***
	(0.135)	(0.120)	(0.116)	(0.118)
Post × Pilot	0.284	0.449***	0.285*	0.432**
	(0.194)	(0.173)	(0.173)	(0.169)
Observations	7725	7725	7725	7725
C. Placebo treatment (assigned randomly by state)				
Post 2002	-3.588***	-3.217***	-68.469***	-58.406***
	(0.210)	(0.235)	(4.364)	(7.784)
Placebo	0.901	0.787	0.904	0.782
	(0.779)	(0.719)	(0.778)	(0.718)
Post × Placebo	-0.353	-0.439	-0.355	-0.444
	(0.446)	(0.481)	(0.446)	(0.473)
Observations	15369	15369	15369	15369
County controls	N	Y	N	Y
Time controls	N	N	Y	Y

Notes: Standard errors clustered at the state level given in parentheses for Panels A and C. Following Bertrand et al. (2004), with small group sizes standard errors are bootstrapped (999 repetitions).
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

(ASDs) with a treatment state in them. Panel B restricts the control group to be only non-pilot neighboring states. Panel C creates a placebo group by random number generation at the state-level, and then assigning all states with a random number greater than the average random value to be treatment. We expect that Panels A and B should confirm our results, which they largely do. In our main specification (Column 4), we find positive and significant effects on the difference-in-difference coefficient for both sample restrictions. Restricting the control group to only include neighboring states weakens the effect, which does suggest that there may be some regional drivers of our main effect.

We expect that the results of the placebo test, which randomly assigns treatment to each state, should be insignificant. We construct our placebo by first assigning a random number on the uniform distribution from zero to one to each state. We then take the average of all random numbers, and assign states to treatment or control based on whether a state's random value is greater than (treatment) or less than (control) the average. This results in roughly 46% of the sample being assigned to treatment. We then re-run the regression given in Equation 3.3 using the placebo group instead of the set of pilot states. We find a negative and insignificant effect of treatment when applying the placebo, confirming our expectations that we should not observe a positive and significant effect when treatment is assigned randomly.

Table 3.4.
Difference-in-difference results on net exit to total operators

	Net exit in 0 to 9 years	Net exit in 5 plus years	Total net exit
Post 2002	-105.077***	73.669***	-31.408***
	(5.148)	(9.178)	(11.373)
Pilot state	0.035	0.560	0.595
	(0.408)	(1.134)	(1.485)
Post × Pilot	0.347	-1.683	-1.336
	(0.539)	(1.813)	(2.298)
Observations	15367	15367	15367
County controls	Y	Y	Y
Time controls	Y	Y	Y

Notes: Standard errors clustered at the state level given in parentheses.

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

Net exit results are given in Table 3.4. There are no significant effects on the difference-in-difference variable. Net exit in the first nine years has a small positive effect from treatment, while net exit for five plus years and all together has a negative effect. Since the LCGP focuses on land contract sales, it's primary effect on exit rates would be through making new farmers more secure in their finances. Since the LCGP would most likely effect long-term viability, and even in that case the effect is likely to be small, it is not surprising that it does not have a profound effect on exit rates. Since there remain no significant effects even under the event study, it is likely that the LCGP impacts entry but not exit.

Table 3.5 presents results on the treatment coefficient for both the fully-dynamic (Equation 3.4) and the static (Equation 3.5) event studies. Panel A presents the results from the fully-dynamic study. The full set of coefficients from three census waves before treatment until treatment are given in Figure 3.8. Both event study specifications include time and county-level fixed effects and cluster standard errors at the state level. The coefficient on treatment is positive and significant in both. As expected from the static model, which replicates the difference-in-difference model but including two waves of treatment, presents a similar (albeit slightly smaller) effect from exposure to the Land Contract Guarantee Program. The fully-dynamic study has a larger effect, but as can be seen in Figure 3.8 the effect may represent a trend towards increased entry in the run-up to treatment. As discussed above, it is likely that using the full data horizon poses challenges given the expansion of incentives in the 2014 Farm Bill. We consider this again below when we present second-stage results using the event study estimates.

Table 3.5.
First-stage event study results for share of total operators new using 2017 data

	(1)	(2)
A. Fully dynamic event study		
Treatment	5.018***	4.490***
	(1.370)	(1.245)
B. Static event study		
Treatment	1.033***	0.896***
	(0.291)	(0.292)
Observations	18446	18446
County controls	N	Y

Notes: Standard errors clustered at the state level given in parentheses. All regressions include time and county fixed effects.
*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

3.4.2 Second-stage results

Table 3.6 presents second-stage results on the full set of outcomes we examine for all four of our specifications: the main difference-in-difference using share of total operators with less than five years experience, the difference-in-difference using share of total operators with less than ten years experience, and the two event studies. For

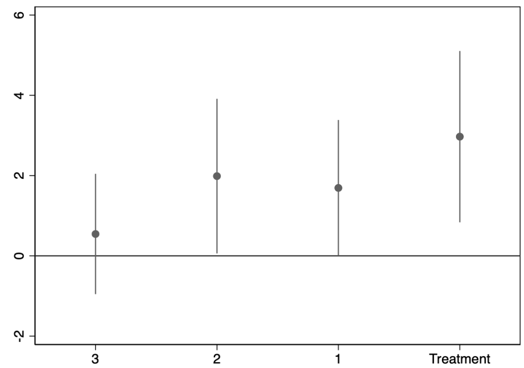


Figure 3.8. Event study coefficients from three census waves before treatment

second-stage regressions we likewise compare a set of methods for calculating standard errors, except that we cluster at the county-level (the same level as our fixed effects), as we are focusing on farm operation decisions that will likely be influenced by county-level space and economies. In general, coefficients from second-stage regressions are more likely to be found significant if standard errors are not clustered. Table 3.8 in the Appendix provides the same set of second-stage results using estimated shares from the robustness checks in the first-stage.

Panel A from Table 3.6 presents the main set of results we are interested in for this paper. These coefficients are the result of using predicted values for share of total operators that are new, estimated using Equation 3.3, in our set of second-stage regressions given by Equation 3.6. As discussed above, we find that counties with higher shares of new farmers have lower participation in federal conservation programs, both as a fraction of total acres in a county and as an absolute value. We are not surprised that farmer entry incentivized by the LCGP would target farmers that are not interested in converting their land to Conservation Reserve Program (CRP) land, given the fact that they must meet regular payments on their guaranteed purchases that likely exceed CRP payments. However, when we consider the full set of years using the event study results, we do find increased participation in conservation programs among new farmers. This may be a result of additional incentives for new and beginning farmers developed in the 2008 Farm Bill and funded and expanded in the 2014 bill Sureshwaran and Ritchie (2011).

Under our main results, we also find lower machinery assets per operation and total in county (assets are expressed in million dollars). We are likewise not surprised

Table 3.6.
Second-stage results

	% Conserved	Total Conserved	Total Machinery	Machinery per operation	Tractors	% tractors new	New tractors per operation	Fertilizer expense per operation	% Insured	Rent per acre
A. Main results										
Estimated % total operators new	-0.498***	-1476.470***	-2.898***	-0.007***	14.133***	-0.567***	0.009**	-5.200***	-19.877*	0.069
	(0.038)	(93.240)	(0.250)	(0.000)	(2.829)	(0.198)	(0.005)	(0.417)	(10.305)	(0.228)
B. Share < 10 years experience										
Estimated % of total operators new	-0.313***	-1784.446***	1.081	-0.005***	-3.995	-0.570***	0.002	-3.778***	-12.599	0.950
	(0.060)	(173.452)	(0.719)	(0.001)	(5.511)	(0.145)	(0.003)	(0.665)	(7.833)	(0.630)
C. Fully dynamic event study										
Estimated % of total operators new	0.328***	632.564***	8.822***	0.013***	-24.182***	0.472***	0.006***	9.775***	-21.494***	6.334***
	(0.051)	(122.610)	(0.394)	(0.000)	(3.066)	(0.080)	(0.002)	(0.565)	(5.895)	(0.330)
D. Static event study										
Estimated % total operators new	0.565***	1486.926***	11.511***	0.016***	-45.288***	-0.618***	-0.007**	10.700***	-17.872***	5.484***
	(0.048)	(111.298)	(0.432)	(0.000)	(4.532)	(0.136)	(0.003)	(0.612)	(4.581)	(0.297)
Observations	10383	12736	15328	15328	15286	12045	12049	12110	8269	11982

Notes: Standard errors clustered at the county level given in parentheses. All regressions include county-varying and time-varying controls. All specifications include county-level fixed effects.

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

that counties with higher shares of new farmers generally have less intensive capital, however the effect is relatively small (a reduction of around 3% of the average asset size). Surprisingly and potentially in conflict with this previous result, we find a larger number of total tractors in counties with higher shares of new farmers. At the same time, the fraction of tractors in a county that are new purchases (< five years old) declines. We are hesitant, however, to over-interpret this result. In general, the share of new farmers has a small and insignificant effect on the count of new tractors and the total tractors per operation (coefficients not reported), although it does have a positive (but small) effect on the number of new tractors per operation. These results may simply suggest noise on capital investment, but it is also possible that new farmers are not necessarily investing in new tractors.

Additionally, we find lower expenditures on fertilizer per acre of cropland. Of particular interest in this result is whether new farmers are more likely to practice organic or less intensive applications of inputs. Unfortunately, with the level of data we currently have we cannot further investigate this question[7]. That said, it is of note that counties with more new farmers have 12% lower levels of a crucial input use. The reversal of this sign under the full event study approach, however, again points to the possible effect of specific types of incentive programs. Future research must better distinguish between incentive types when comparing the types of farmers and the effect of new farmer behavior on county aggregates.

A broadly consistent result is that counties with higher shares of new farmers experience significantly lower use of federal crop insurance programs. This result is significant at only the 10% level in our main result, but is strengthened when including the full data horizon. The fact that counties with higher shares of new farmers experience lower levels of insurance coverage confirms our expectations, since participation is notably difficult for new farmers who are unaware of all available options or lack the institutional connections to take full advantage of federal crop insurance programs. We are surprised, however, that including the 2017 data does not abate this reduction, since an important policy introduction in the 2014 Farm Bill was increased support for crop insurance for new and beginning farmers. We plan to more closely examine this by using improved data from the Agricultural Resource Management Survey that includes more farm-level data on crop insurance decisions.

Finally, we find no significant effect on rent per acre in our main set of results. The minimal effect on the rental market from an intervention in the land sale market

[7]Data on organic operations was collected only in the 2007 Agricultural Census wave.

is surprising, but likely reflects the limited set of qualified buyers and sellers under the LCGP. Since rental markets are crucial in agriculture, the size and scope of land rental likely exceeds the subset of the land-for-sale market that the LCGP influences. Similar to other results, the strengthening effects under the event study designs, including the static event study, suggests that as the program was rolled out nationwide, the impact of participation in the land-sales market may have had rental effects.

3.5 Conclusion

This paper investigates the impact of new and beginning farmers on farm operation decisions at the county level. While estimating the impact of shares of new farmers on farm operation measures is difficult due to simultaneity and omitted variable bias, we use variation in the timing of a policy designed to encourage the sale of land to new and beginning farmers to estimate county-level shares of operators who are new. The Land Contract Guarantee Pilot Program, which was launched in 2002 in six states, provided loan guarantees to new and beginning farmers when buying land from existing farmers. The guarantees were meant to make it easier for new and beginning farmers to own land for agriculture, focusing on one of the major issues preventing entry into the sector (Mishra et al., 2009).

Using a differences-in-differences model, we estimate that the effect of the LCGP on pilot states was a 0.99% increase in the share of new farmers. While this value is low in absolute terms, it represents roughly 8.4% of the national average and 10% of the average in pilot states. In addition, it represents an abatement of around 2% of the secular trend in declining shares of new operators from 1992 to 2012. The sign and significance of these results are confirmed in two robustness checks that narrow our control sample to more similar areas, and a placebo test for treatment finds insignificant results.

An event study model that takes into account the nation-wide roll out of the LCGP in 2012 finds a 4.49% increase in the share of new operators in the periods following treatment (i.e. availability of the LCGP). Since the LCGP requires qualifications that may not be fully reflected in the share of operators who are new and given the limited number of available guarantees in the pilot program, all of our results should be seen as intent-to-treat effects. Nevertheless, we expect that our measure of new farmers should bias our results towards insignificance (since our shares are upper bounds on the actually-treated group). The effect of the LCGP still largely appears to be positive, which is noteworthy in a period of steadily declining shares of new farmer entry.

Using the results from our first-stage, we estimate the impact of county-level shares of new farmers on a variety of aggregate measures of farm operational decisions. While the results from this second stage should be viewed with some skepticism, since county-level aggregates of farm operation decisions are noisy, we find a range of expected and significant results. In particular, we find that counties with higher shares of new operators have lower participation in federal conservation programs and federal crop insurance programs. The former is particularly expected given the limitations of CRP payments to meet land sales contracts. In addition, we find that counties with higher shares of new farmers have less intensive capitalization at the farm level. This confirms results from Mishra et al. (2009), Kropp and Katchova (2011), and Ahearn (2011). We also find lower expenditures per acre of cropland on fertilizer, although absent more information on production systems we hesitate to interpret this result. Nevertheless, it points in the direction of possible variation in type of farm production method for new farmers relative to existing farmers.

Apropos that question, a future extension of this research must take into account a broader range of policy interventions. Federal incentives for new and beginning farmers include additional financial supports, including authorization in 2008 and appropriations in 2014 for the sale of Conservation Reserve Program land by retiring farmers to new farmers. That policy includes support for new farmers to begin Environmental Quality Incentives Program (EQIP) projects on CRP land prior to expiration of the conservation agreement. That means that new and beginning farmers can begin land improvements on conservation land where existing farmers cannot, increasing the value of CRP land and increasing the likelihood that new and beginning farmers own CRP land—a result similar to what we find when we consider the post-2012 census wave.

In addition to financial incentives, the USDA offers risk management and educational benefits for new and beginning farmers. The 2014 Farm Bill authorized the Risk Management Agency (RMA) to offer farmers with less than five years experience as a principal operator favorable terms in crop insurance, including reduced fees. Given the lower share of acreage under federal crop insurance programs in counties with high shares of new operators, this policy seems especially relevant. Multiple educational programs have been put in place, many of which involve grant applications from regional extension offices. Future research must include all of these incentives, taking advantage of variation in timing and intensity of exposure to these policies. This paper provides a roadwork for doing so, using the Land Contract Guarantee Program pilot study as an example.

3.6 Appendix

Table 3.7.
Summary statistics for control and pilot states, only including neighboring states in control

	Control	Pilot	Total
Total operators	746.62	883.78	783.81
% of operators < 5 yrs. exp.	11.08%	9.98%	10.78%
% of operators < 10 yrs. exp.	26.16%	23.94%	25.56%
Total acres cropland	108,071.75	138,876.48	116,425.01
Total acres pasture	87,497.79	48,794.81	77,002.77
% acres under conservation programs	3.67%	4.55%	3.93%
Total acres under conservation programs	12,735.80	14,610.09	13,290.71
Total machinery assets (Million $'s)	74.72	98.35	81.14
Machinery assets per operator (Million $'s)	96,410.04	111,570.53	100,525.06
Total tractors in county	1,619.77	2,093.87	1,748.46
% tractors new	10.29%	8.79%	9.88%
Fertilizer expenses per acre ($)	43.31	45.73	43.97
% acres under federal crop insurance	43.79%	51.01%	45.85%
Rent per acre ($'s)	22.28	33.83	25.46

Table 3.8.
Second-stage results using robustness checks from first stage

	% Conserved	Total Conserved	Total Machinery	Machinery per operation	Tractors	% tractors new	New tractors per operation	Fertilizer expense per operation	% Insured	Rent per acre
A. Limited sample										
Estimated % of total operators new	-0.523***	-1434.585***	-2.263***	-0.007***	8.439***	-0.071	0.018***	-5.606***	-18.034*	0.062
	(0.043)	(102.140)	(0.280)	(0.000)	(2.936)	(0.200)	(0.005)	(0.455)	(9.760)	(0.271)
Observations	9821	12039	14471	14471	14439	11372	11376	11456	7818	11316
B. Only neighbors										
Estimated % of total operators new	0.530***	-773.307***	-3.945***	-0.007***	23.091***	-2.188***	-0.043***	6.035***	0.338	6.730**
	(0.147)	(74.679)	(0.260)	(0.000)	(3.414)	(0.299)	(0.007)	(2.183)	(1.186)	(2.798)
Observations	5420	6690	7710	7710	7693	6042	6043	6086	4193	5992
C. Placebo										
Estimated % of total operators new	-0.480***	-1609.859***	-5.799***	-0.009***	28.924***	-1.373***	-0.004	-6.222***	-20.636	-1.296***
	(0.039)	(92.225)	(0.215)	(0.000)	(3.002)	(0.397)	(0.008)	(0.434)	(13.342)	(0.236)
Observations	10383	12736	15328	15328	15286	12045	12049	12110	8269	11982

Notes: Standard errors clustered at the county level given in parentheses. All regressions include county-varying and time-varying controls. All specifications include county-level fixed effects.

*, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively.

BIBLIOGRAPHY

Abrams, D. and Fackler, R. (2018). To plea or not to plea: Evidence from north carolina. Working paper.

Abrams, D., Galbiati, R., Henry, E., and Philippe, A. (2018). Judicial delegation. Working paper.

Abrams, D. S., Bertrand, M., and Mullainathan, S. (2012). Do judges vary in their treatment of race? *The Journal of Legal Studies*, 41(2):347–383.

Ahearn, M. (2011). Potential challenges for beginning farmers and ranchers. *Choices*, 26(2).

Ahearn, M. and Newton, D. (2009). Beginning farmers and ranchers. Working paper.

Bar-Gill, O. and Ayal, O. G. (2006). Plea bargains only for the guilty. *Journal of Law and Economics*, XLIX:353–364.

Bellin, J. (2018). Reassessing prosecutorial power through the lens of mass incarceration. *Michigan Law Review*, 116.

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

Bibas, S. (2004). Plea bargaining outside the shadow of trial. *Harvard Law Review*, 117(8):2464–2547.

Bigelow, D., Borchers, A., and Hubbs, T. (2016). Us farmland ownership, tenure and transfer. Working paper.

Bindler, A. and Hjalmarsson, R. (2017). The persistence of the criminal justice gender gap: Evidence from 200 years of judicial decisions. Working Paper.

Boehlje, M. (1973). The entry-growth-exit process in agriculture. *Southern Journal of Agricultural Economics*, 5(1):23–36.

Borusyak, K. and Jaravel, X. (2017). Revisting event study designs, with an application to the estimation of the marginal propensity to consume. Working paper.

Boyd, C. L. (2016). Representation on the courts? the effects of trial judges' sex and race. *Political Research Quarterly*, 69(4):788–799.

Boylan, R. T. (2012). The effect of punishment severity on plea bargaining. *Journal of Law and Economics*, 55:565–591.

Boylan, R. T. and Long, C. X. (2005). Salaries, plea rates, and the career objectives of federal prosecutors. *Journal of Law and Economics*, XLVIII:627–651.

Carr, J. B. and McClain, W. (2019). Evidence on court system bias from strategic judge assignment. Working paper.

Dahl, G. B., Kostol, A. R., and Mogstad, M. (2014). Family welfare cultures. *The Quarterly Journal fo Economics*, 129(4):1711–1752.

Depew, B., Eren, O., and Mocan, N. (2017). Judges, juveniles, and in-group bias. *The Journal of Law and Economics*, 60(2):209–239.

Devers, L. (2006). Plea and charge bargaining: Research summary. Technical Report 2008-F08151, Bureau of Justice Assistance, U.S. Department of Justice, Washington, D.C.

Dobbie, W., Goldin, J., and Yang, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2):201–240.

Dobbie, W. and Song, J. (2015). Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *American Economic Review*, 105:1272–1311.

Frandsen, B. R., Lefgren, L. J., and Leslie, E. C. (2019). Judging judge fixed effects. Working paper.

French, E. and Song, J. (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, 6(2):291–337.

Gale, H. (2003). Age specific patterns of exit and entry in u.s. farming, 1978-1997. *Review of Agricultural Economics*, 25(1):168–186.

Golin, E. (1995). Solving the problem of gender and racial bias in administrative adjudication. *Columbia Law Review*, 95(6):1532–1567.

Gonzalez-Uribe, J. and Reyes, S. (2019). Identifying and impacting "gazelles": Evidence from business accelerators. Working paper.

Green, D. P. and Winik, D. (2010). Using random judge assignments to estimate the effects of incarceration and probation on recidivism among drug offenders. *Criminology*, 48:357–387.

Grossman, G., Gazal-Ayal, O., Pimentel, S. D., and Weinstein, J. M. (2016). Descriptive representation and judicial outcomes in multiethnic societies. *American Journal of Political Science*, 60(1):44–69.

Grossman, G. M. and Katz, M. L. (1983). Plea bargaining and social welfare. *American Economic Review*, 73:749–757.

Hernandez-Julian, R. and Tomic, A. (2006). How elected judges respond to the racial composition of their constituencies. *Journal of Legal, Ethical and Regulatory Issues*, 17(2).

Johnson, B. D. (2014). Judges on trial: A reexamination of judicial race and gender effects across modes of conviction. *Criminal Justice Policy Review*, 25(2):159–184.

Jr., J. J. D., Graves, J. A., Gruber, J., and Kleiner, S. A. (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy*, 123(1):170–214.

Katchova, A. and Ahearn, M. (2016). Dynamics of farmland ownership and leasing: Implcations for young and beginning farmers. *Applied Economic Perspectives and Policy*, 38(2):334–350.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review*, 96(3):863–876.

Kropp, J. and Katchova, A. (2011). The effect of direct payments on liquidity and repayment capacity for beginning farmers. *Agricultural Finance Review*, 71(3):347–365.

Kutateladze, B. L., Andiloro, N. R., and Johnson, B. D. (2016). Opening pandora's box: How does defendant race influence plea bargaining. *Justice Quarterly*, 33:398–426.

Landes, W. M. (1971). An economic analysis of the courts. *The Journal of Law & Economics*, 14:61–107.

Lim, C. S., Silveira, B. S., and Snyder, J. M. (2016). Do judges' characteristics matter? ethnicity, gender, and partisanship in texas state trial courts. *American Law and Economics Review*, 18(2):302–357.

Maestas, N., Mullen, K. J., and Strand, A. (2013). Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *American Economic Review*, 103(5):1797–1829.

Metcalfe, C. and Chiricos, T. (2017). Race, plea, and charge reduction: An assessment of racial disparities in the plea process. *Justice Quarterly*, pages 1–31.

Mishra, A., Wilson, C., and Williams, R. (2009). Factors affecting financial performance of new and beginning farmers. *Agricultural Finance Review*, 69(2):160–179.

Mnookin, R. H. and Kornhauser, L. (1979). Bargaining in the shadow of the law: The case of divorce. *The Yale Law Journal*, 88(5):950–997.

Norris, S. (2019). Examiner inconsistency: Evidence from refugee appeals. Working paper.

Norris, S., Pecenco, M., and Weaver, J. (2019). The effects of parental and sibling incarceration: Evidence from ohio. Working paper.

Piehl, A. M. and Bushway, S. D. (2007). Measuring and explaining charge bargaining. *Journal of Quantitative Criminology*, 23:105–125.

Price, J. and Wolfers, J. (2010). Racial discrimination among nba referees. *The Quarterly journal of economics*, 125(4):1859–1887.

Quintanar, S. (2017). Man vs. machine: An investigation of speeding ticket disparities based on gender and race. *Journal of Applied Economics*, 20(1).

Reinganum, J. F. (1988). Plea bargaining and prosecutorial discretion. *American Economic Review*, 78:713–728.

Rokicki, S., Cohen, J., Fink, G., Salomon, J. A., and Landrum, M. B. (2018). Inference with difference-in-differences with a small number of groups. *Medical Care*, 56(1):97–105.

Rubin, J., Jr., P. R. D., and Grine, A. A. (2013). Indigent defender manual.

Schanzenbach, M. (2005). Racial and sex disparities in prison sentences: The effect of district-level judicial demographics. *The Journal of Legal Studies*, 34(1):57–92.

Schanzenbach, M. M. and Tiller, E. (2005). Strategic judging under the united states sentencing guidelines: Instrument choice theory and evidence. In *American Law & Economics Association Annual Meetings*, page 12. bepress.

Sorensen, T. A., Sarnikar, S., and Oaxaca, R. L. (2010). Do you receive a lighter prison sentence because you are a woman or a white? an economic analysis of the federal criminal sentencing guidelines. *The BE Journal of Economic Analysis & Policy*, 14(1):1–54.

Starr, S. B. and Rehavi, M. M. (2012). Racial disparity in the criminal justice process: Prosecutors, judges, and the effects of united states v. booker. Working Paper.

Stuntz, W. J. (2004). Plea bargaining and criminal law's disappearing shadow. *Harvard Law Review*, 117:2548–2569.

Sureshwaran, S. and Ritchie, S. (2011). U.s. farm bill resources and programs for beginning farmers. *Choices*, 26(2).

Tomic, A. and Hakes, J. K. (2008). Case dismissed: Police discretion and racial differences in dismissals of felony charges. *American law and economics review*, 10(1):110–141.

Uggen, C. (2000). Work as a turning point in the life course of criminals: A duration model of age, employment, and recidivism. *American Sociological Review*, 67:529–546.

West, J. (2017). Racial bias in police investigations. Working Paper.

William McClain

339 Charles Street
Rockville, MD 20850
Phone: (765) 421-1080

Website: www.willmcclain.com
Email: `wbmccl@gmail.com`
United States Citizen, Born in NY

Education

Ph.D. candidate Agricultural Economics, Purdue University (Expected completion in 2019).
Dissertation: *Incentives and organization in policy*
Fields and topics: Applied microeconomics; public economics; law and economics; contract theory; policy analysis and program evaluation

M.S. Agricultural Economics, Universität Hohenheim, Stuttgart, Germany 2015.

B.B.A. Accounting, Woodbury University, Burbank, California 2012.

Research

Projected climate change impacts on Indiana's energy demand and supply, 2019. Leigh Raymond, Doug Gotham, William McClain, Sayanta Mukherjee, Roshanak Nateghi, Paul V. Preckel, Peter J. Schubert, Shweta Singh, and Elizabeth Wachs. *Climatic Change*, https://doi.org/10.1007/s10584-018-2299-7.

You are not alone: Social capital and risk exposure in rural Ethiopia, 2016. Tesfamichael Wossen, Salvatore Di Falco, Thomas Berger, and William McClain. *Food Security* 8, 799–813.

A pathway forwards for the social capital metaphor, 2016. William McClain. *Review of Social Economy* 74:2, 109-128.

Contingent liabilities after BP Oil spill, 2012. Ashley Burrowes, John Karayan, and William McClain. *Chartered Accountants Journal of New Zealand*, 90:5.

Working paper

***Evidence On Court System Bias From Strategic Judge Assignment*, Jillian B. Carr and William McClain**: Bias is a known issue in the U.S. courts system, and in this paper, we present evidence of strategic behavior by prosecutors in case design, scheduling and plea offers. We use administrative data from the North Carolina criminal justice system and the rotation of elected judges, and we find that case timing, speed of adjudication and sentences respond to the judge rotation in ways that indicate strategic assignment of judges to cases. This has consequences both for case outcomes and for the use of judicial rotation as a quasi-experimental research design.

Presentations

Evidence of court system bias from strategic judge assignment, Southern Economic Association Annual Meetings, Washington DC (2018)

Recent teaching and research experience

Research assistant, Spring 2019-Present; Supervisor: Dr. Bhagyashree Katare
Working with Dr. Katare cleaning and analyzing time series, panel, and survey data on nutrition and health

Teaching assistant, Fall 2018-Present
AGEC 217: Economics; Supervisor: Dr. Lawrence DeBoer
Introductory undergraduate class in macroeconomics

Sole Instructor of AGEC 327: *Food and Agribusiness Marketing*, Spring 2018
Created and evaluated quizzes, in-class exercises, homeworks and exams
Topics covered: Marketing strategy and issues in global food and agricultural marketing.

Sole Instructor of AGEC 516: *Mathematical Tools for Agricultural and Applied Economics*, Fall 2017
Designed and implemented lectures for masters-level mathematics course; wrote notes, homeworks, and exams. Topics covered: Algebra, multivariate differential calculus, matrix algebra, optimization

Academic Tutor, Brees Academic Performance Center, Purdue University, 2016-2017
Tutored student athletes in AGEC 217 (Economics), ECON 210 (Principles of Economics), ECON 251 (Microeconomics), ECON 360 (Econometrics)

Research assistant, 2015-2017; Supervisor: Dr. Paul Preckel
Worked on a team examining the effects of climate change on Indiana's energy system.

Teaching assistant, Fall 2016
The Macroeconomic, Trade, and Policy Environment of the Food System; Supervisor: Dr. Philip Abbott
Online masters-level class for working professionals as part of joint MS-MBA with Indiana University - Bloomington.

Additional information

Fellowships and awards: Frederick N. Andrews Fellowship, Purdue University 2015-2017; Purdue Climate Change Research Center Incentive Award, 2015

Languages and skills: English (native), German (intermediate); Stata, R, Matlab, GAMS, Latex, Microsoft Office

Refereeing and service: Reviewer at *Review of Social Economy*, Graduate Student Organization president and academic chair, Seminar Committee, College of Agriculture Graduate Advisory Committee, Student organization advisor

References

Steven Yu-Ping Wu
Associate Professor
Department of Agricultural Economics
Purdue University
(765) 494-4299
sywu@purdue.edu

Jillian Carr
Assistant Professor of Economics
Krannert School of Management
Purdue University
(765) 496-0288
carr56@purdue.edu

Gerald E. Shively
Professor
Department of Agricultural Economics
Purdue University
(765) 494-4218
shivelyg@purdue.edu

Last updated: October 9, 2019