ESSAYS IN APPLIED MICROECONOMICS

A Dissertation

by

BRITTANY RENAE STREET

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Mark Hoekstra
Committee Members,	Jason Lindo
	Steven Puller
	Lori Taylor
Head of Department,	Timothy Gronberg

May 2019

Major Subject: Economics

ABSTRACT

In an era of increasing data availability, it is possible to better understand the world around us and improve policy with empirical evidence. The difficulty is in distinguishing causal effects from other confounding factors in quantitative analysis. The purpose of each of the following essays is to uncover causal effects using natural-experiments and various methodologies along with detailed administrative data. The first essay studies the extent to which investments in physical education during middle school can improve student health and achievement. The second essay examines ways in which jurors may be systematically biased, specifically focusing on the interaction between the gender of the defendant and the jury composition. The final essay studies how individuals respond to improved job opportunities with respect to their criminal behavior. Collectively these studies contribute to discussions on health, education, the criminal justice system, discrimination, stimulus programs, and hydraulic fracturing.

DEDICATION

To my family, Brad, Deanna, Lynnsey, Brisben, Mary, Abigail, and Samuel: this would not have been possible without your love, support, and belief in me.

ACKNOWLEDGMENTS

I can not overstate my gratitude to my advisor, Mark Hoekstra. The time spent going over results, learning to identify credible research designs and spot problems, and instilling confidence in my abilities was instrumental in shaping this dissertation and my career. A special thanks to Steven Puller, my committee member and coauthor on projects not in this dissertation, for making site visits to both Ghana and North Dakota possible. The time and energy invested in our projects and my professional development is greatly appreciated. Additionally, thank you to my committee members Jason Lindo, who first introduced me to causal inference, and Lori Taylor for offering valuable feedback and support during my time at A&M. Finally, I would like to acknowledge Jennifer Doleac for mentorship in the final stages of the Ph.D. program and job search.

I would also like to thank my fellow Aggies, including coauthor Analisa Packham, co-advisees Abigail Peralta, CarlyWill Sloan, Meradee Tangvatcharapong, and Adam Bestenbostel, office-mates, and many others. The camaraderie and mentorship were and are invaluable. I can not thank Brian Prescott enough for the immense effort he devoted as an undergraduate research assistant. I would also like to thank the staff in the department, namely Chelsi Bass, Lynn Drake, Teri Tenalio, Mary Owens, Ludim Garcia, and Kurt Felpel, for all the help behind the scenes. Moreover, I am grateful for the support from the Private Enterprise Research Center at Texas A&M.

Importantly, I would like to acknowledge the organizations that provided the data for the research in this dissertation. The Education Research Center and Texas Education Agency provided access to school- and student-level data used in Section 2. Hillsborough and Palm Beach County Clerk's office graciously provided records and institutional details for Section 3, with special thanks to Angela Gary, Keith Buss, and Tara Ramos. The final chapter was made possible thanks to the State of North Dakota Judicial Branch, especially Jeff Stillwell, and DrillingInfo.

Last and certainly not least, I would like to thank my family for their continued support. In particular, my dad for instilling a strong work ethic, my mom for keeping me sane, and Sam for keeping life on track.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Mark Hoekstra, primary advisor, as well as Jason Lindo and Steven Puller of the Department of Economics and Professor Lori Taylor of the Bush School of Government and Public Service.

Data for each chapter was acquired and analyzed by the student. Section 1 is joint work with Professor Analisa Packham and Section 2 is joint work with Professor Mark Hoekstra. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University including a diversity fellowship, the Gail Frey Monson Memorial Fellowship, the Summer Time Advancement in Research Award, and a fellowship with the Private Enterprise Research Center.

NOMENCLATURE

API	Application Programming Interface
BEA	Bureau of Economic Analysis
BMI	Body Mass Index
CDC	Center for Disease Control
ERC	Education Research Center
FDR	False Discovery Rate
FE	Fixed Effect
GEEG	Govenor's Educator Excellence Grant
HFZ	Healthy Fitness Zone
IRS	Internal Revenue Service
LEOKA	Law Enforcement Officers Killed and Assaulted
OLS	Ordinary Least Squares
PE	Physical Education
PERC	Private Enterprise Research Center
RDD	Regression Discontinuity Design
STAAR	State of Texas Assessment of Academic Readiness
TAKS	Texas Assessment of Knowledge and Skills
TEA	Texas Education Agency
TEEG	Texas Educator Excellence Grant

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

This dissertation, and my research more broadly, centers on social and policy relevant questions in the field of applied microeconomics spanning topics in crime, education, and health. Much of my research agenda is driven by understanding the effects of policy or identifying ways in which policy could be improved. Methodologically, I identify causal effects using quasi-experiments coupled with detailed administrative data.

In my first chapter, Analisa Packham and I examine the impact of investments in physical education (PE) by analyzing a four-year, $37 million program in Texas that mandated daily PE classes for middle-school students. Despite the fact that several public health agencies have promoted physical activity in schools as a way to combat growing rates of childhood obesity, little is known about how such initiatives affect student outcomes. We estimate the causal effects of PE on student fitness, achievement, and classroom behavior by exploiting a discontinuity in program eligibility to analyze the impact of TFN on over 350,000 students. We find no evidence that such investments lead to healthy changes in body composition, higher overall fitness levels, or improvements in student test scores. On the contrary, we find some suggestive evidence that physical education has adverse consequences for middle-school students, including more classroom misbehavior and reduced school attendance.

In my second chapter, I along with Mark Hoekstra examine the extent to which criminal conviction rates are affected by the similarity in gender of the defendant and jury. To identify effects, we exploit the random variation in both the assignment to jury pools and the ordering of potential jurors. We do so using detailed administrative data on the juror selection process and trial proceedings for two large counties in Florida. Results indicate that own-gender juries result in significantly lower conviction rates on drug charges, though we find no evidence of effects for other charges. Estimates indicate that a one standard deviation increase in expected own-gender jurors (10 percentage points) results in an 18 percentage point reduction in conviction rates on drug charges, which is highly significant even after adjusting for multiple comparisons. This re-

sults in a 13 percentage point decline in the likelihood of being sentenced to at least some jail time. These findings highlight how drawing an opposite-gender jury can impose significant costs on defendants, and demonstrate that own-gender bias can occur even in settings where the importance of being impartial is actively pressed on participants.

Finally, my third chapter studies how individuals respond to improved economic conditions with respect to their criminal behavior. Economic theory suggests crime should decrease as economic opportunities increase the returns to legal employment. However, there are well-documented cases where crime increases in response to areas becoming more prosperous. This paper addresses this puzzle by examining the effects on crime only for residents already living in the area prior to the economic boom. This approach isolates the effect of local economic opportunity from the effect of changing composition due to in-migration during these periods. To identify effects, I exploit within- and across-county variation in exposure to hydraulic fracturing activities in North Dakota using administrative individual-level data on residents, mineral lease records, and criminal charges. Results indicate that the start of economic expansion —as signaled by the signing of leases —leads to a 22 percent reduction in criminal cases filed. Effects are smaller once the fracking boom escalates during the more labor-intensive period. This is consistent with improved economic opportunity reducing crime.

2. THE EFFECTS OF INVESTMENTS IN PHYSICAL EDUCATION ON STUDENT HEALTH AND ACHIEVEMENT

2.1 Introduction

In the United States, the rate of childhood obesity has more than quadrupled in the past thirty years (Centers for Disease Control and Prevention, 2016). One in three children are at risk of becoming overweight or obese, and among children of lower socioeconomic status, the risk is even higher (Centers for Disease Control and Prevention, 2016; Let's Move, 2016).

From a public policy perspective, policies that seek to target the inputs to obesity, like food and exercise, can reduce negative externalities imposed by higher health care costs in the long run (Cutler et al., 2003; Finkelstein et al., 2003; 2009). Given that children between the ages of 5-18 spend approximately 40 hours a week at school and may eat several meals there, a natural policy solution to address childhood obesity and increase total social welfare is to encourage children to form healthy habits at school. The purpose of this paper is to analyze the effects of one such initiative, Texas Fitness Now, on student health and academic performance.

Due to the concern of rising health risks and costs of obesity in recent years, federal and state agencies have created new guidelines and implemented numerous programs to encourage physical activity. Recently, medical authorities including the Institute of Medicine, American Heart Association, and the American Academy of Pediatrics, have endorsed curricula that consist of at least 30 minutes of daily physical activity a day as a way to reduce obesity and overweight (Institute of Medicine, 2013; Pate and O'Neill, 2008; Wilson, 2017). Despite these recommendations, schools may not provide enough opportunities for students to meet this standard during the school day, due to resource or time constraints. Only 3.8% of elementary schools, 7.9% of middle schools, and 2.1% of high schools provide daily physical education (Centers for Disease Control and Prevention, 2007).

Although physical education (PE) interventions are continually recommended by medical pro-

fessionals as a strategy to increase physical activity and reduce childhood obesity, the results of such policies have been mixed. A literature review by Guerra et al. (2013) reports that only 1 out of 11 published studies that use randomized control trials to evaluate PE programs estimate significant reductions in body mass index (BMI). None find effects on body weight. And while a handful of studies document that increasing PE time can reduce obesity for young, elementary-school children (Centers for Disease Control and Prevention, 2016; Cawley et al., 2013; Waters et al., 2011; Datar and Sturm, 2004), there is less evidence to suggest that such programs are effective at reducing BMI for middle-school or high-school students (Cawley et al., 2007; Wang et al., 2003; von Hippel and Bradbury, 2015; Knaus et al., 2018).

Separate from the effects on health, PE proponents argue that increasing physical activity yields large academic benefits by improving cognition, focus, and memory. There is a growing body of research implying that this may indeed be the case.[1] In a recent report, the CDC describes analyses that link school-based physical activity, including physical education, to academic behaviors such as cognitive skills, academic attitudes, attendance, and achievement, and provides suggestive evidence of a positive relationship between physical activity and academic performance. (Centers for Disease Control and Prevention, 2010).[2] Moreover, studies evaluating increases PE time in schools appear to offer some affirmation that such programs can improve student outcomes (Tremarche et al., 2007; Carlson et al., 2008).[3]

That being said, one concern is that increasing PE requirements takes away important instructional time, which could lead to less learning and poorer student outcomes. In a review of 7 quasi-experimental studies, which focus on academic outcomes for students up to grade 6, Trudeau and Shephard (2008) finds that physical activity can be added to school curriculum without hindering student achievement. Dills et al. (2011) similarly explores this hypothesis and finds no statistically

[1] For evidence on the relatinoship between physical activity and cognition, see, for example, Tomporowsk et al. (2008).

[2] Out of the 43 studies, nearly all estimates testing the relationship between academic performance and physical activity are positive (98.5%), and approximately half are statistically significant.

[3] Specifically, Tremarche et al. (2007) estimates the effects of a randomized control trial, and concludes that students in an elementary school with more PE time had higher reading test scores. Carlson et al. (2008) uses Early Childhood Longitudinal Study and finds that increasing PE time raises test scores for girls.

significant or economically significant impact of weekly PE on test scores for elementary-aged children, suggesting that PE at worst has no effect on academic achievement.

Based on the above research, we would expect that policies targeting physical activity have the ability to positively affect student behavior and performance, implying that there may be some scope for school-level services to play an even larger role. However, nearly all of the literature to date focuses on elementary-aged children, while little to no evidence exists on the effects of PE on middle-school students. Accordingly, a fundamental policy question remains unanswered: how much can PE affect adolescent fitness and health, and how much do these programs translate to changes in attendance, disciplinary action, and academic performance?

To answer this question, we present new evidence on the effects of physical education requirements and contribute to a growing literature on how policies can address childhood obesity and student achievement. In particular, we estimate a model that exploits a discontinuity in eligibility criteria for Texas Fitness Now (TFN), a four-year physical education grant program targeting low-income students with the aim of improving overall health and well-being. Program-eligible schools included campuses teaching grades 6, 7, and/or 8 with a large majority of economically disadvantaged students. Participating schools received funds contingent on the agreement that they: (i) spent funds on new athletic equipment or services related to PE and (ii) ensured that students attend PE classes for 30 minutes each school day.

The Texas Education Agency (TEA) has since pointed to the positive improvements in fitness and body composition as evidence of the program's success; however, the fact that fitness scores were increasing in each subsequent year of the program suggests that other factors probably contributed to the average increase observed across some Texas schools (Texas Education Agency, 2011).[4]

Similarly, von Hippel and Bradbury (2015), uses a fixed effects model instrumenting for program participation over time, and estimates that TFN improved some measures of fitness for some

[4] In particular, the TEA compared the year-to-year differences in test scores in grantee schools only. They report that TFN led to statistically significant increases of 3.6-6.2 percentage points in aerobic capacity, trunk lift, upper body strength and endurance, and body composition between 2007 and 2009 (Texas Education Agency, 2011).

groups, although they find no effects on BMI.[5] However, the authors model estimates by gender as well as groups of years of the program separately, making both the overall average effects and local average effects of the program difficult to distinguish, and they do not provide any support for their identifying assumption, casting doubt on the validity of the research design.

This research addresses these shortcomings and builds on the existing literature in a number of ways. First, we employ a regression discontinuity design, using the eligibility criteria directly, to compare otherwise similar students across the TFN eligibility threshold. Under the plausible assumption that other determinants of student fitness and performance are smooth across the school grant-qualifying cutoff, this research design allows us to compare outcomes of students in schools just below the eligibility threshold to students just above the threshold. In doing so, we are able to provide evidence that any changes in student outcomes are a result of the program, and not an artifact of school selection or other unobservable characteristics. Below we provide evidence in support of this assumption showing schools eligible for physical education grant funding were similar to schools just below the cutoff in terms of size, other financial resources and student composition. Second, we use individual-level administrative data to study student outcomes, which constitutes an improvement on school-level data since it additionally contains information on student raw test scores, attendance and disciplinary behavior. Moreover, the granular nature of these data allows us to test for compositional changes and detect student attrition.

We find that Texas Fitness Now did not improve physical fitness, including overall body mass index (BMI), body fat, aerobic capacity, or strength and flexibility. However, we show that TFN was effective at reducing the number of obese students, implying that such interventions may be most effective for high-risk students.

Using individual-level data on student academic outcomes for Texas middle schoolers, we estimate no effect of the program on student achievement. Conversely, we present suggestive evidence that compulsory PE classes reduce attendance rates and increase incidents of disruption and mis-

[5] In particular, von Hippel and Bradbury (2015) finds that effects were greatest in measures of strength, and greater for girls than for boys, although they report no statistically significant effects on shoulder flexibility. They find that both boys and girls in high-poverty middle schools could complete more pushups and a faster shuttle run. Girls could also complete more curl-ups, a higher trunk lift, and had a better sit and reach.

behavior. These findings imply that interventions encouraging daily physical activity have the potential to negatively impact students if adolescents have a strong aversion to physical education.

Given the existing literature documenting the beneficial effects of physical education on elementary-aged students, these findings may be somewhat surprising. However, there are several potential explanations why 12-14 year olds respond differently than young children to physical activity initiatives. For instance, middle-school students may have already formed lifetime exercise and eating habits, and therefore are more obstinate than elementary-aged children to abandon unhealthy habits. Another explanation is that for economically disadvantaged and/or overweight teens, PE class may serve as a class period where students that struggle with aerobic exercises experience bullying and teasing.[6] Finally, since middle-schoolers are less energetic, physical education could make teenagers more tired than younger students, which may contribute to more distractions or misbehavior in the classroom. Below, we explore these possibilities in an effort to shed light on the more comprehensive effects of PE requirements for middle-school students.

The rest of this paper is organized as follows. Section 2 describes the Texas Fitness Now program in more detail. Section 3 describes the data and empirical strategy. Section 4 presents the main results. Section 5 provides a discussion of the main results and potential mechanisms before concluding.

2.2 Background on Texas Fitness Now

The Texas Fitness Now (TFN) program was, at the time of initiation, the second-largest physical activity grant program in the US.[7] From 2007-2011, with the goal of curtailing childhood obesity and Type II diabetes, the State of Texas allocated $37 million to the poorest Texas middle schools to be spent on physical education and activity. Although nutrition was included as a facet of the program, TFN primarily focused on increasing funds and requirements for physical edu-

[6] In particular, students report being bullied more in middle school than at any other point during their academic career. Over 22 percent of middle schoolers experience bullying at least once per week, as compared to 11 percent of high schoolers, and these effects are largest in low-income schools (National Center for Education Statistics, 2018).

[7] For reference, the largest grant program is the ongoing yearly Carol M. White Physical Education Program, which allocated 72.6 million in grants for physical education to 149 entities in 2007 (U.S. Department of Education, 2013).

cation.[8] Schools that accepted the funding were required to have students participate in physical education classes for at least 30 minutes per day or 225 minutes every two weeks.[9] To ensure compliance, applicants detailed how they would feasibly incorporate more PE classes into their curriculum, and participants were required to conduct fitness assessments twice per school year for evaluation. While no data exists on how individual schools allot time for PE, during the course of the program, over 1/3 of participating campuses reported having difficulty finding time to fit more PE classes into the curriculum, indicating that the program's time constraints were binding for many schools (TEA 2011).[10]

Table A.1 displays the total amount of grant funding distributed in each year of the program, as well as how eligibility changed over time. Schools serving 6th, 7th, and 8th grade students were eligible to participate in the grant program if in the previous school year they had reported having at least 75% economically disadvantaged students, although this cutoff was extended to include schools with 60% economically disadvantaged students in 2009 and 2010.[11] Participating schools received an average of $10,000 to improve their physical education programs by purchasing equipment such as stop watches, pedometers, jump ropes, and free weights, as well as by adding more PE classes and hiring coaches and fitness instructors.[12]

The State of Texas required that the schools use the grant money as a supplement and not as a replacement for other academic programs. For example, TFN funds could not be spent on

[8] Only 7% of schools reported spending some money on nutrition initiatives (Texas Education Agency, 2011).

[9] While Texas maintains baseline PE requirements for middle schools, students in grades 6-8 only need to participate in daily physical activity for 4 out of 6 semesters. "Physical activity" is defined at the district-level, but in many cases may include extra-curricular activities, such as marching band or cheerleading, although these activities would not meet PE requirements under TFN guidelines.

[10] One of the main limitations of our data is that we are unable to speak directly to how schools chose to reallocate timing for PE courses. After speaking to a few administrators, we note that the most popular route that schools took to implement the program was to reallocate elective course blocks to physical education. Therefore, it does not appear that many schools reduced time for math and reading as a result of the program. Instead, students would spend a semester in PE instead of courses such as art, theater, computer programming, or choir. Unfortunately, since our data do not contain any information on student courseload, we are unable to observe changes in student schedules.

[11] Economically disadvantaged students are indicated as students that: (i) are eligible for free or reduced-price meals, (ii) are from a family with an annual income at or below the poverty line, (iii) are eligible for public assistance, and/or (iv) received any need-based financial assistance.

[12] Funding for eligible schools was determined by a fixed amount ($1,500) plus a proportional amount ranging from $11-$32 per 6th-8th grade student, depending on the school year.

athletics or construction projects. During the first three years of the program, over 60% of schools spent money on traditional equipment, while, on average, 15% and 24% of schools added staff and classes, respectively. Nearly all of the participating schools reported that after receiving grant money they were able to provide opportunities for students to participate in physical activity at least 30 minutes a day or 225 minutes per two weeks (Texas Education Agency, 2011).

2.3 Empirical Approach

This section describes the data and approach we use to estimate the causal effects of the Texas Fitness Now program on student health, fitness, test scores, discipline, and attendance.

2.3.1 Data

Data on fitness outcomes are from a statewide testing assessment for physical fitness, known as the FITNESSGRAM© test. These data are collected by health educators in the spring of each school year and are available from school years spanning 2007-2013. Given confidentiality concerns, FITNESSGRAM© data are available only by school, gender, and grade in the spring of each academic year. Notably, this limitation of the data means that we are unable to examine differences in physical ability across race, ethnicity, or fitness level. Students are tested in six main areas: body composition, aerobic capacity, upper body strength and endurance, abdominal strength and endurance, flexibility, and trunk extensor strength and flexibility.

Given their age and gender, results are measured relative to a range of acceptable scores for each test, known as the “healthy fitness zone" (HFZ). The HFZ is intended to reflect the level of fitness needed for good health. Students are not informed of the HFZ cutoff intentionally as a way to motivate them to perform their best. Since a majority of students are able to achieve their HFZ for all tests, any failures signal a need for more frequent exercise.

See Figure A.10 for an official chart of healthy fitness zones for each fitness test, by age and gender. For body composition, HFZ levels represent a healthy weight. However, FITNESSGRAM© additionally indicates where students “need improvement"; these upper ranges correspond to overweight or obesity. Otherwise, HFZ ranges are intended to represent a typical level of fitness by age

and gender. For example, according to Figure A.10, a 13-year-old girl would need to complete 18 curl-ups and 7 push-ups to pass the corresponding fitness tests.

We additionally analyze effects of physical education on student performance and classroom behavior at the student level using data from the Education Research Center at UT-Austin. These data include student demographics, including economically disadvantaged status, as well as raw test scores, attendance, and disciplinary behavior for the full population of students enrolled in a Texas public school from school years spanning 2006-2011. One of the main advantages of these data is that we are able to examine effects on individual test scores and attendance as well as student discipline and suspensions, which are unavailable in the publicly available, school-level data. Additionally, these data allow us to analyze heterogeneous effects by subgroups, such as grade or gender. Moreover, we are able to rule out compositional changes in student cohorts across schools due to the policy, which allows us to test for student attrition. To study effects of TFN on middle-school students, we limit our sample to the population of Texas students in grades 6, 7, and/or 8 for the school years of the program (2007-2008 to 2010-2011).

Summary statistics for student characteristics and outcomes are presented in Table A.7. Testing performance rates for reading and mathematics standardized tests, known as the Texas Assessment of Knowledge and Skills (TAKS) tests, are defined as whether a student met or exceeded the passing standard set by the state in the corresponding school year. Mean passing rates for math and reading TAKS tests range from 71-83 percent.

We categorize student-level disciplinary action into three main outcomes: total number of incidents, whether or not the student ever misbehaved, and total days of suspension in a given school year. On average, a student is reprimanded for one disciplinary incident per school year; however, only 27% of students misbehave in a given school year.

Attendance outcomes are based on mean student attendance rates for the entire school year. Student attendance rates are calculated by dividing the total number of days a student was present by the total number of eligible school days. As shown in Table A.7, attendance in any given year

is very high (above 95%).[13] That said, any measured changes in attendance rates are likely to be small.

2.3.2 Identification strategy

Measuring the causal effects of Texas Fitness Now presents many challenges. For example, eligible schools self-select into TFN, and may additionally receive funding from other government programs, such as Title I, or the National School Lunch Program. Schools with the most motivated and ambitious faculty may therefore be those that choose to participate, and any estimated positive effects will overstate the benefits of the program.

To overcome such challenges and estimate effects of Texas Fitness Now, we use a regression discontinuity design (RDD). This strategy exploits the cutoff in program eligibility, the percent of economically disadvantaged students, to identify the causal effects of increased physical education requirements. Our approach is motivated by the idea that characteristics related to behaviors and outcomes of interest are likely to vary smoothly through this threshold. Thus, any discontinuity in fitness, test performance, discipline, or attendance can be reasonably attributed to the change in the physical education curricula. We operationalize this identification strategy by estimating:

$$y_s = \theta EDcutoff + f(EDpct_s) + \lambda_t + \psi_g + \eta_s \tag{2.1}$$

where y_s is an average measure of student fitness, for school s or academic performance, attendance, and discipline outcomes for student s, f is a function of the percent of economically disadvantaged students for school s in school year 2006-2007, and $EDcutoff$ is a binary indicator for whether a school s meets the first year eligibility cutoff, as listed in Table A.1. Because the program spans four years and multiple grade levels, we additionally include year fixed effects, λ_t in all specifications, and in some specifications we include grade fixed effects, ψ_g.[14] We control

[13] Notably, days suspended do not count as an absence.

[14] As suggested by Lee and Card (2008) we use data on baseline covariates, including student race, ethnicity, gender, and economically disadvantaged status and school characteristics, such as student population, only to test the validity of the RD design, although below we additionally discuss results from models in which we add student-level demographic controls.

for the percent of economically disadvantaged students, normalized to zero, (running variable) linearly and allow it to vary on either side of the eligibility cutoff. Following Lee and Card (2008), we present standard errors that are clustered on the running variable, although we note that our estimates are not sensitive to this choice.[15]

Although, in practice, school eligibility for TFN was reevaluated each year, we use only the first year of eligibility criteria (the percent of economically disadvantaged students in 2006-2007) to define a school's position relative to the qualifying cutoffs in each year. In holding each school's eligibility constant across all years, we prevent the possibility of strategic schools manipulating their position across the threshold over time. Estimates based on this approach should yield more conservative estimates than those that depend on the yearly definition of treatment; however, we note that estimates for our preferred specifications are statistically similar for all outcomes using either approach.[16]

Our preferred specifications show estimates from all four school years, 2007-2008 to 2010-2011, using a bandwidth of 15 on either side of the cutoff. Given that the cutoff was expanded from 75 to 60 halfway through the program, we consider this bandwidth to be the largest possible range that exploits the variation in eligibility criteria, although we present estimates for a range of bandwidths in Figure A.13.[17]

The identifying assumption for this research design is that characteristics related to outcomes of interest vary smoothly through the treatment threshold. Since eligibility for Texas Fitness Now

[15] We cluster on the running variable since the percent of economically disadvantaged students is rounded to the nearest tenth of a percentage point, although estimates are robust to clustering at the school level. Specifically, estimates on overall fitness levels and test scores remain statistically insignificant, while we estimate a statistically significant decrease in attendance rates and increase for all discipline outcomes at the full bandwidth of 15.

[16] We additionally note that the percent of economically disadvantaged students in a given school is highly correlated across years, and schools are unable to choose which students attend.

[17] A one-sided bandwidth greater than 15 would contain schools which may have been treated every year in the program. However, these schools, which contain a large proportion of economically disadvantaged students, may not be an appropriate comparison group for schools that fall just short of program participation after the expansion in eligibility. For example, a school with 80 percent economically disadvantaged students would be eligible for the program in the first two years, given the cutoff of 75, and we would effectively be comparing these students to those in schools with 70 percent economically disadvantaged students. However, when eligibility is expanded to 60, if we included this school in our analysis, we would estimate a local average treatment effect that effectively compares these same students to those in schools with 40 percent economically disadvantaged students. Therefore, we exploit the 15 percentage point expansion in program eligibility to estimate the local average treatment effects for students in schools that would not have already been treated prior to this criteria change.

is based on a school's previous year's percent of economically disadvantaged students, this feature helps to assuage concerns that the identification assumption may not hold.[18] Additionally, because schools likely do not have control over which students move into or out of their school district in any given year, manipulation of the running variable is unlikely. We test for this possibility, as suggested by McCrary (2008), in several ways. First, we confine the percent of economically disadvantaged students by school to be that of the first year program criteria, which eliminates the possibility for schools to move across the cutoff in subsequent years. In doing so, we estimate intent-to-treat effects, which will likely understate the true effects of the program.[19] Second, we test for discontinuities of several school characteristics, such as race, gender and ethnic composition as well as total number of students fitness tested across the eligibility threshold to address the possibility that unqualified schools close to the cutoff were systematically different than those that barely qualified for funding. Third, we show that the percent of economically disadvantaged students does not exhibit a discontinuity at either the 60% or 75% cutoff, which provides some support for the notion that the State of Texas chose these cutoffs arbitrarily and schools were not able to manipulate around them. Fourth, we present evidence that student selection into or out of program-eligible schools is not driving our results by providing estimates of the number of schools that a student attends during the sample period. Fifth, we estimate Equation 1 for all outcome variables using pre-period data from 2006 to show that any estimated effects for 2007-2011 are a result of the program, and not an existing feature of the data.

With any education-based school reform, it is important to consider whether there are additional grants available for schools that meet this same cutoff, which may lead to additional treatments that affect academic outcomes but are unrelated to physical fitness. Indeed, Title I funding, which is set aside for schools with at least 40% of economically disadvantaged students is a major source

[18] Schools are required to report the percent of economically disadvantaged students in October of the current school year. The Texas Comptroller announced original TFN eligibility cutoffs in June 2007, which suggests that schools were unaware of the threshold when reporting students statistics to the TEA in the previous year.

[19] Effects are similar when allowing for school eligibility status to vary across treatment years; we estimate no statistically significant effect on test scores, which corresponds to Columns 3 and 6 in Table A.3, an increase in disciplinary incidents of 0.07, which corresponds to Column 3 in Table A.4, and a decrease in attendance rates of 0.002, which corresponds to Column 3 in Table A.5.

of school funding and provides an average of $630,000 dollars to Texas schools each year. We provide evidence in Figure A.11 that this cutoff is not sharp, as many schools with small shares of economically disadvantaged students still receive Title I funds.[20]

Texas did initiate two performance incentive programs in 2006, the Governor's Educator Excellence Grant (GEEG) and the Texas Educator Excellence Grant (TEEG), as a way to increase quality of education through higher pay for school personnel and professional development. Although one component of eligibility for funds was based on the number of economically disadvantaged students, schools also were required to display acceptable testing performance. Due to this additional requirement, over 200 schools below our treatment threshold participated in these two grant programs, indicating that additional grant funding was continuous through the TFN eligibility cutoff. Furthermore, few middle schools participated in these two programs; over half of TEEG and GEEG funds went to Texas high schools, which are not included in our sample. Importantly, neither program used the 60 or 75 percent economically disadvantaged as a funding criteria.[21] Finally, we note that to the extent that these other school resources improve academic performance and/or attendance, any negative findings of TFN on performance would be understated.[22]

2.4 Main Results

2.4.1 Effects of Texas Fitness Now eligibility on funding

Figure A.1 presents the estimate for the main measure of TFN participation: total grant money awarded. Here we present residual means plots (accounting for year and grade fixed effects) using 3 percentage point bins as well as the respective discontinuity estimates from Equation 1. In all figures the running variable (the percent of economically disadvantaged students) is normalized to

[20] Similarly, there exist community standards for the National School Lunch Program, in which a school with many economically disadvantaged students are eligible for funds to provide lunch to all students. However, to participate in this program, schools receive funds on a phase-in rate, starting at the 42.5 percent economically disadvantaged student cutoff. Therefore, we would not observe a discontinuity of funds at the 60 or 75 percent cutoffs due to this program. Moreover, the Community Eligibility Provision was rolled out in Texas in 2013, which should mitigate any concern that discontinuities in school lunch funding is driving our results.

[21] Additionally, we find no evidence of an existing discontinuity at the TFN eligibility threshold on total school funding and total operational expenses ($p > 0.8$).

[22] To our knowledge, there are no other grants that utilized the same economically disadvantaged cutoffs during our sample period.

zero due to the fact that this cutoff changed in 2009.

As shown in Figure A.1, we estimate statistically significant discontinuities in funding based on the eligibility cutoff, although we note that this criteria is not sharp.[23] Specifically, schools that met the eligibility criteria received, on average, approximately $10,600 in TFN funding, which corresponds to approximately $15 per student. We note that, while $15 per student does not seem like a large intervention, this represents about a 6% increase in total per pupil instructional spending and is 17 times the average Texas middle-school PE budget. Moreover, cost-benefit analyses of similar physical activity interventions estimate spending requirements of only approximately $4 per student to increase physical activity to 30 minutes per day and subsequently decrease obesity by 0.02 BMI units over two years (Barrett et al., 2015). Furthermore, the $15 per student average includes some schools that did not receive any TFN funding; however, take-up of the program was high with approximately 88-95 percent of eligible Texas schools both applying and receiving the grant in a given year.

2.4.2 Effects of Texas Fitness Now on fitness

Since the intent of TFN was to improve fitness outcomes and reduce obesity for middle-school students, in this section we present estimated discontinuities for body composition and physical fitness outcomes, including measured tests for BMI, body fat, aerobic activity, strength and flexibility. Importantly, these data are only available by school, grade, and gender, and are not obtainable at the individual level.

TFN participation stipulated that students attend PE class every day for at least 30 minutes. Since a majority of schools in Texas do not have requirements for the length of PE class, and many schools do not require students to attend PE for all three years of middle school, this was likely a noticeable change in curricula for many students (CDC, 2007). Indeed, a large majority of schools (82-87 percent) reported being able to restructure curriculum to meet this requirement

[23] Although eligibility was intended to limit funding only to middle schools, eligibility was also extended to alternative schools with any grade level. About 6 percent of Texas schools that received funding did not contain students in 6th, 7th, or 8th grades. We do not include any of these schools in our analyses.

(Texas Education Agency, 2011).[24,25]

We first show effects of TFN on body composition. Importantly, the data do not include information on student-level BMI calculations; we only have information on the percent of students with a healthy BMI, students that are at-risk, or are overweight, and students that have high-risk, or are obsese. Figure A.2 shows residuals means plots for the percent of students with a healthy body-mass index using 3 percentage point bins.[26] This figure shows some support for the notion that TFN was ineffective at reducing BMI for low-income students.

In Table A.2, we display the corresponding point estimates. Each column is a separate regression, and each regression uses data for all 6th, 7th, and 8th graders in Texas from school years 2007-2008 to 2010-2011. In Column 1, we first estimate the optimal (bias-corrected) bandwidth and polynomial order, as suggested by Calonico et al. (2016). This procedure specifies one-sided optimal bandwidths ranging from 5.3 to 13.5 and first-order polynomials for all outcomes. In Column 2 we adopt a bandwidth of 12, for comparison, while in Column 3 we display estimates using a bandwidth of 15, which is the full bandwidth using the expansion in eligibility criteria in 2009.

As shown in Column 1, we estimate that TFN led to approximately a 2.2 percentage point *reduction* in the percent of students with a healthy BMI. This could be due to several reasons. For example, if students are working out more, they could be counteracting the effects of physical activity by eating more calories. Or, perhaps students are more tired and therefore less likely to play sports at home or participate in after-school activities. Another possibility is that students face bullying or hardship in the locker room and become discouraged or give up trying to lose weight. However, estimates in Columns 2 and 3 are statistically insignificant, indicating that the program likely had no effect on student BMI. Based on the estimates in Table A.2, we can rule out effects of a 0.03 percentage point increase of students with a healthy BMI, or a 0.46% increase.[27]

[24] In Texas middle schools that have a physical education requirement, there is no requirement for everyday physical activity. Students are required to attend PE class the equivalent of 225 minutes per two weeks or 30 minutes per day for four semesters, but may choose which semesters to participate.

[25] Since the TEA does not maintain records on block schedule schools, we are unable to test differences between students with an A/B class schedule and students with 7-8 class periods every day.

[26] Notably, Texas schools that use FITNESSGRAM© as a measure of physical fitness have flexibility to choose which measure of body composition to report- over 75 percent report BMI.

[27] Another possibility is that, given the metrics of “healthy", “at-risk", and “high-risk", it is possible that TFN had an

We also note that there is heterogeneity across student preferences for physical fitness; therefore, it may be more informative to analyze the effects of daily PE classes on students that are overweight versus students that are obese. We present estimates and their corresponding 95% confidence intervals across a range of bandwidths for these respective groups of students in Figure A.3. Across all bandwidths, we find that the number of obese students decreased as a result of the program, which implies that although the intervention was ineffective at helping students reach a healthy BMI overall, and may have *increased* weight for some students, such policies may be able to help the heaviest individuals lose weight.

Importantly, although it may be difficult for school-mandated PE classes to affect BMI in an economically meaningful way, we may expect that overall physical fitness levels would improve. To observe effects of TFN on a broad measure of fitness, we construct a school-level variable for the average number of fitness tests passed and present these estimates in Columns 4-6 of Table A.2. These tests include aerobic activity, strength and flexibility and do not include measures of BMI. Estimates shown in Table A.2 are precise enough to rule out even small increases in the number of tests passed (1.9 percent), implying that TFN did not marginally increase fitness levels, on average.

Finally, we test for more specific indicators of physical fitness, measured by FITNESSGRAM© tests, and present results in Figure A.12. These tests include aerobic capacity, strength, and flexibility.[28,29] Estimates for all fitness outcomes are statistically insignificant and indicate that TFN had little to no effect on aerobic capacity, strength, or flexibility among middle-school students. These estimates are similar in magnitude across all bandwidths.[30]

average, positive effect on BMI, but this effect was not large enough to move students into or out of the various categories.

[28] FITNESSGRAM© provides opportunities for schools to test strength and flexibility in a variety of ways. These tests include curl-ups, trunk lift, 90 degree push-ups, pull-ups, flexed arm hang, sit and reach, and shoulder stretch. See `http://pyfp.org/doc/fitnessgram/fg-07-muscular.pdf` for a description the objectives, scoring, and instructions for each test.

[29] In testing aerobic activity, schools have the option to complete the pacer test or have students complete a mile run without stopping. Nearly 75% of schools opt for the pacer test over the mile run. The pacer test, also known as the progressive aerobic cardiovascular endurance run, is a multistage shuttle run designed to test endurance and aerobic capacity by requiring students to run across a 20-meter space at a specified and increasing pace, making the test increasingly more difficult as time progresses.

[30] While many reports have pointed to the positive outcomes for fitness, especially for young girls, we find no major differential effects of TFN on physical fitness outcomes by gender (TEA, 2011). See Table A.8 for effects of TFN on body composition, aerobic capacity, and strength and flexibility for girls.

Contrary to von Hippel and Bradbury (2015), we find little evidence that TFN improved student fitness levels. However, we note that it is possible that TFN failed to encourage students that were already relatively healthy to marginally pass more fitness tests, but was able to target those students with the worst levels of physical fitness. We also note that, while, on average, TFN did not reduce the number of overweight students, if daily PE classes increase physical activity for sedentary adolescents, students may still gain other, unobserved, independent health benefits (Institute of Medicine, 2012).

2.4.3 Effects of Texas Fitness Now on test scores

Given the potential for changes in PE curricula to affect student focus and achievement, we now examine the effects on academic outcomes. Specifically, the State of Texas measures academic performance for grades 3-12 based on passing rates for reading and mathematics on standardized tests, known as the Texas Assessment of Knowledge and Skills (TAKS) tests.[31] TAKS subject tests measure knowledge on the state-mandated curriculum objectives and consist of multiple-choice questions scored by a computer. Scores are scaled and the passing score levels change slightly from year to year depending on the test's level of difficulty. According to data from the Texas Education Agency on state testing, TAKS attendance and completion is 99-100% for all years during the sample period.

For students, the TAKS test represents a high-stakes test that they must sit for once a year in the spring. If a student does not pass either the math or reading exams at the end of the 8th grade year, they are not permitted to advance to high school. If a student fails either exam in the 6th or 7th grade, they may advance grades, but are required to take additional remedial courses to catch up to the knowledge level of their peers. We focus our analyses on exams that students must take every year, namely math and reading.[32]

[31] From 2012-14 the TAKS test was phased out, as Texas switched to the State of Texas Assessments of Academic Readiness (STARR) test. Therefore, we do not analyze any longer-term effects of TFN on school years 2011-2012 or 2012-2013, after the program had ended due to concerns of comparability.

[32] While some middle-school students are required to additionally test for writing, social studies and science in some years, we limit the analysis to reading and mathematics TAKS scores, given that all students take these tests each year from 3rd-11th grade. When estimating effects for these alternative subject tests, we find no evidence that PE investments affect the percent of students that pass.

In Figure A.4 we present evidence that TFN did little to improve student performance, as measured by TAKS passing rates and raw test scores. Mirroring these findings, Table A.3 displays estimates on passing rates for math and reading TAKS scores from a baseline specification derived from Equation 1, controlling for grade and year fixed effects.[33] Passing grades are determined by the Texas Education Agency, and are measured by the number of questions answered correctly compared to the passing standard set by the state in the corresponding year. We additionally show estimates for whether students received a "commended" recognition, a distinction of high achievement that only 20-33% of students receive in a given year, and the number of questions the student answered correctly, i.e. the raw TAKS scores.[34]

Although the TEA reports that daily PE requirements have the potential to increase test scores, we find little evidence to support this finding (TEA, 2011). Estimates across all columns of Table A.3 indicate statistically insignificant effects of TFN on both math and reading scores. These effects are consistent across specifications and are precise enough to rule out effects on math and reading passing rates of 0.56% and 0.36% percent or larger, respectively.[35] Therefore, our findings suggest that investments in physical education do not negatively (or positively) affect overall student performance, which is consistent with previous studies on adolescent physical activity.[36]

Since TFN was geared towards helping economically disadvantaged students, and since we may expect fitness interventions to affect students differently by gender, we additionally analyze

[33] In all specifications using individual-level data we control for year and grade fixed effects, although we note that our results are not sensitive to the inclusion of grade controls.

[34] Specifically, the State of Texas designates a students' score to be "commended" if they score at least 2100 out of 2400 scaled points.

[35] These effects are relatively small when compared to effects found using first-order academic interventions. For comparison, assignment to smaller class sizes in the well-known Tennessee STAR experiment in grades K-3 increased student test scores in grades 6-8 by 3.6-6.0 percentile points (Schanzenbach, 2007). Similarly, students in grades 4-8 lotteried into New York City charter schools gained 12 and 9 percent of a standard deviation each year on math and English test scores, respectively (Hoxby et al., 2009). Our estimates suggest that students spending up to 2.5 hours more per week in PE gain less than 0.4 percentile points in math, with smaller effects for reading, or less than 0.9 percent of a standard deviation increase.

[36] For other studies that analyze the effects of physical education interventions on student performance, see Dills et al. (2011), and Cawley et al. (2013), von Hippel and Bradbury (2015). In particular, Dills et al. (2011) estimates a value-added model and finds that weekly PE classes have no statistically significant or economically significant impact on test scores for elementary-aged children. Cawley et al. (2013) uses the Early Childhood Longitudinal Study, Kindergarten Cohort and instruments for child PE time, according to state policies. They find no evidence of spillovers of PE on test scores for elementary school children. von Hippel and Bradbury (2015) uses school-level data to study TFN and finds no effect of the grant program on academic achievement.

how TFN affected test scores for students across these subgroups in Table A.9.[37] Effects for females and economically disadvantaged students are positive and statistically similar to estimates of the overall sample, suggesting that these estimates are not driven by one particular group.[38]

2.4.4 Effects of Texas Fitness Now on disciplinary action

Although there is little evidence to suggest that mandatory PE classes affect student health and fitness, such initiatives may affect student behavior in a number of ways. First, it's possible that PE classes encourage restless students to expel nervous energy, allowing them to focus more on coursework, and be less disruptive throughout the day. However, if students become more tired throughout the day due to the increase in physical activity and/or have strong preferences against such classes, we would expect an increase in misbehavior. In Figure A.5 and Table A.4, we provide some evidence to suggest the latter.

Before discussing statistical evidence of TFN on disciplinary action, we first present visual evidence that mandatory PE requirements affect student behavior in the classroom. Figure A.5 displays the effect of TFN on the total number of student disciplinary incidents, proportion of student offenders, and total days suspended. Each figure shows large, positive discontinuities at the eligibility cutoff. Overall, the set of results in Figure A.5 indicate that daily PE requirements lead to more recorded instances of student misbehavior.[39]

Table A.4 shows additional estimates from regressions with smaller bandwidths. Models with optimally-chosen bandwidths (Columns 1, 4, and 7) as well as models with a one-sided bandwidth

[37] We also provide estimates for the effect of TFN on test scores by grade in Table A.10. We find no statistically different effects of daily PE requirements by grade level.

[38] We could also measure the effects of TFN on test performance, discipline, and attendance by race and ethnicity, however, we do not include these subsamples in our main analyses for two reasons. First, we are unable to examine effects of TFN on fitness by race and ethnicity. Second, we estimate a small and statistically significant discontinuity at the 10% level for some outcomes one year before the program ($p \geq 0.09$), although we do not find such a discontinuity in aggregate outcomes. These effects yield some concerns that the RD model may be misspecified when looking at some subgroups, thus we omit any analysis by race and ethnicity throughout the paper.

[39] Arguably, we may expect schools that hired more staff to be able to report more disciplinary incidents due to increases in monitoring. Unfortunately, we do not have data on school-level expenditures from the TFN grant funding and are unable to speak to this mechanism directly. However, we acknowledge that the increases in disciplinary action that we observe in the data are not borne entirely by a small population of schools, which lends some evidence to the argument that these effects are at least partially student-driven. Moreover, only 7 percent of TFN schools added staff from 2008-2010, indicating that monitoring is unlikely to be responsible our results.

of 12 (Columns 2, 5, and 8) yield statistically similar but insignificant estimates. Therefore, despite the proposition that PE classes incite student focus and good behavior, we show no evidence that TFN *reduced* classroom disruptions. However, we do present some evidence in Columns 3, 6, and 9 that daily compulsory PE requirements may actually *increase* instances of classroom misbehavior. Estimates from a model with the full bandwidth indicate that TFN resulted in a statistically significant increase of 0.15 incidents for each student, on average, which corresponds to an increase in disciplinary action of about 15.6%, or 73 per school year.

Notably, this measure could represent an increase on either the intramargin or inframargin; that is, either students that were already likely to misbehave did so more frequently, or there were more instances of new offenders. We investigate the extent to which one of these effects is driving the total effect in Columns 4-6 of Table A.4, which presents the proportion of total students that caused a classroom disruption. In Column 6, we estimate that TFN increased the proportion of misbehaving students by 0.02, or 7.4%. Therefore, we report suggestive evidence that daily PE classes for middle-school students may not only lead to more disciplinary action but also encourage more students to act out.

Finally, as a way to analyze the intensity of student misbehavior, we investigate how many days students were suspended as a result of disciplinary infractions and present results in Columns 7-9. Although estimates are not statistically significant across all bandwidths, estimates in Column 9 indicate that TFN increased the number of days suspended by 23.7 percent. In terms of class time, this corresponds to about 0.84 fewer days of traditional coursework for misbehaving students in TFN-eligible schools, as compared to students in the non-eligible middle schools.[40]

One explanation of these findings is that mandatory PE classes increase bullying in school. Although the ERC student-level data does not contain information on *where* the instances of disciplinary action occurred, it is possible that more frequent interaction in the locker room leads to more teasing and fighting throughout the school day. Given that nearly all cases of US school

[40] We additionally provide estimates of disciplinary action by grade in Table A.10 as well as by gender and economically disadvantaged status in Table A.9. While estimates are larger in magnitude for 8th graders, estimates are not statistically different at the 1% level across grades, gender, or economic status. Similar to findings in Table A.4, estimates are less precise at smaller bandwidths.

infractions occur in the classroom, (e.g. 60 percent of major offenses and over 70 percent of minor offenses (Gion et al., 2014)), it may also be possible that both classroom *and* locker room bullying increases as a result of more PE days, but the lack of visibility from teachers in gym escapes punishment. These implications are especially worrisome, given that such bullying can be counterproductive to the goals of physical education programs, as children who are criticized for their physical skills or ostracized in gym class perform worse in school and experience a decrease in physical health and fitness in the long run (Jensen et al., 2013).

2.4.5 Effects of Texas Fitness Now on attendance

If students' preferences for physical education differ from that of other school subjects, increasing PE requirements may affect incentives for student attendance. We test this hypothesis in Figure A.6 and Table A.5. In Table A.5, Column 1 shows estimates from a model based on Equation 1 that uses a MSE-RD estimated optimal bandwidth. Estimates are similar across columns and indicate that TFN did not encourage students to attend school more frequently. Although the baseline attendance rates are high, for some bandwidths we observe a statistically significant decrease in attendance rates for students in TFN-eligible schools as a result of the program. Estimates across Columns 2-3 in Table A.5 indicate that mandatory PE classes reduce attendance for all students by 0.30 percentage points, or 0.31 percent. These findings suggest that, at best, investments in physical education do not cause students to change their decision to come to school; at worst, daily PE mandates could discourage some students from attending class.[41]

In Tables A.9 and A.10, we additionally explore discontinuities in average attendance rates for different student subgroups, including gender, economic status, and grade across the cutoff. We find that effects on attendance are larger for economically disadvantaged students, although effects are not statistically different from the full sample. We find no differential effects by gender (Table A.9) or grade (Table A.10).

These findings suggest that in low-income schools, mandatory PE classes could potentially

[41] Notably, student suspensions do not factor into attendance as an absence. Therefore, it's not the case that the increase in disciplinary action is driving the reduction in attendance rates.

discourage student attendance. Four arguments support this idea: (i) overweight or unathletic students may fear being ostracized or face bullying in the locker room, and would rather skip school than face hardship, (ii) students may fear activities such as running and jumping are too difficult and prefer not to exercise at school, (iii) adolescents concerned for their appearance may not want to look sweaty or untidy during the school day, and/or (iv) middle-school students do not enjoy engaging in movement or physical activity.

It is well-documented that preferences for physical activities and recreation change as students mature. Accordingly, adolescents' overall level of physical activity decreases significantly in 7th and 8th grade at a critical time of physical and cognitive development, especially among girls, with only 17% meeting the daily activity guideline by age 15 (Nader et al., 2008). Given that physical activity after elementary school progressively decreases, the drop in attendance could reflect taste-based preferences for sitting in a classroom over exercising at school (Butt et al., 2011). However, taken with the positive effects of disciplinary action reported in the previous section, TFN may have increased bullying enough to discourage some students from attending school.[42] In either case, to the extent that attendance is crucial for attaining knowledge, paramount for a student's academic success, or is beneficial for emotional or social growth, the effects discussed above are of considerable consequence.

2.4.6 Robustness checks

As discussed in Section 3, we perform a number of robustness checks to provide additional support for the identification assumption. There may be some concerns that schools just above the eligibility cutoff are systematically different than schools just below the cutoff. For example, if schools that participate in TFN have a different composition of students, our findings may be picking up differential behavioral reactions to PE requirements across students. Moreover, if schools receiving TFN funding want to report improved fitness scores as a way to motivate future

[42] We also acknowledge that one plausible alternative explanation is that injuries could result from increased physical exertion that also lead to more student absences. While we cannot directly address this issue using available data, according to conversations with PE coaches at various Texas high schools, injuries in class are not particularly common. Furthermore, the general policy is that injured students with a doctor's note would be allowed to sit on the sidelines and theoretically would not be expected to miss more than one class day due to an injury.

state funding opportunities, coaches may encourage the out-of-shape students to sit out of class on testing days (although this technically violates FITNESSGRAM© rules), which would overstate any positive fitness results in schools just above the eligibility threshold. To test for randomness in the eligibility cutoff, we estimate effects of the percent of economically disadvantaged students in the 2006-2007 school year on the total number of students, the total number of students fitness tested, the percent of female students, the percent of black students, the percent of Hispanic students, and the percent of economically disadvantaged students in our sample and present these results in Figure A.7. Across all outcomes these estimates are statistically insignificant at the 5% level, providing some support that schools on either side of the cutoff are similar on measurable characteristics.[43]

While we do estimate a statistically significant effect at the 10% level of nearly 4 percentage points for the proportion of black students at the cutoff ($t = 1.75$), we note that controlling for demographics yields estimates that are nearly identical. In particular, estimated effects from our preferred specification for attendance and all discipline outcomes are statistically similar at the 1% level when including controls for race and ethnicity. Moreover, we similarly estimate a statistically significant discontinuity of 3 percentage points in the proportion of black students prior to the program's initiation, but do not estimate significant effects for student outcomes in this period, implying that any changes observed in fitness, academic performance, attendance, and discipline after 2006 is a result of the intervention and not racial composition.

In Figure A.8, we additionally test for the density of the running variable, the percent of economically disadvantaged students. To the extent that schools are aware of the eligibility cutoff and can manipulate this threshold, there will be a discontinuity in the number of schools in each bin. Estimates indicate that there is no discontinuity in the number of schools just above and just below the cutoff, suggesting schools did not manipulate the cutoff to receive TFN funding. Similarly, Figure A.14 shows the average number of schools that a single student enrolled in during the four-year sample period to test for student attrition. Estimates are statistically insignificant across all

[43] We also note that, when replicating figures similar to Figure A.13, estimates on all school characteristics are statistically insignificant at the 5% level across all possible bandwidths.

bandwidths, indicating that students did not actively manipulate around the TFN eligibility cutoff.

Next, in Figure A.9, we present evidence that any discontinuities in test scores, disciplinary incidents, and attendance are the result of the program, and not preexisting anomalies in the data. To this end, we replicate our findings from Equation 1, limiting our sample to the school year before the TFN program began, 2006-2007. We estimate no statistically significant discontinuities in test rates, disciplinary incidents, or attendance rates, which provides additional support for the notion that investments in PE programs, and not other factors, are driving our main results.

Finally, we note that since our procedure to determine the optimal bandwidth and polynomial order does not relax our assumption of linear fit and uniform weighting on either side of the cutoff, we have additionally analyzed how our estimates change under different functional forms and show these results in Table A.6. When we impose second- and third-order polynomial fits into our main regression equation, we observe that, while the magnitude of the estimates remain consistent, the significance of the estimates decreases dramatically. Notably, since the choice of polynomial order seems to have a large impact on precision, this may indicate that using higher-order polynomials causes us to overfit the data. Similarly, using triangular kernel weights (Column 4) estimates yield similar effects for attendance as compared to the baseline results in Column 1, and estimates for discipline remain similar in magnitude.

2.5 Discussion and Conclusion

This paper analyzes the effects of increased physical education requirements on student health, fitness, academic performance, and student misbehavior. Using a regression discontinuity approach, we estimate that school-level interventions mandating daily PE classes do not lead to overall improvements in student fitness, including cardiovascular endurance, strength, and flexibility. In particular, although the goal of TFN was to reduce BMI, we show empirically that the program was ineffective at achieving this goal, on average, although we provide some evidence to indicate that TFN was effective at reducing BMI for the most at-risk students.

Moreover, we find that TFN did not lead to positive spillover effects in the classroom, including improvements in math and reading passing rates. However, we present some evidence that daily

PE may be detrimental to student behavior, resulting in increases in disciplinary incidents and reductions in attendance. Given the current recommendations of daily compulsory PE by agencies such as the CDC as well as the US Surgeon General, these findings can better inform policymakers of the effectiveness and potential unintended consequences of such policies for adolescents.

Unfortunately, a limitation of the available data is the inability to accurately test for all the possible mechanisms that explain these results in the classroom. One potential explanation is that requiring students to spend more time in PE class only reduces time spent in other electives, like theater and choir. Alternatively, if students experience diminishing returns to learning, we may expect that as long as the time spent in PE class does not disproportionately take away from one particular academic subject, test performance should be unaffected. In either case, because students are not significantly reducing learning time in math and reading during the day, they perform similarly on standardized tests. Given that, in some cases, we estimate adverse consequences in attendance and disciplinary incidents, the null average effect for test scores seems surprising. One might expect disruptions in class or absences to lead to less learning overall. Although our results point to no effect on student learning, we acknowledge another possibility: athletically inclined students enjoy PE classes and perform better on exams, while those that are most negatively effected by the program perform worse. In this scenario, we would similarly estimate a zero effect on test scores, although we would expect the policy implications to vary based on the composition of students. However, we note that we do not find evidence of such heterogeneous effects across student subgroups of grade, gender, and economic status.

While these explanations are important to consider in terms of student achievement, they do not explain why we observe a decrease in attendance rates and an increase in disciplinary behavior for students at TFN-eligible schools. One mechanism that explains both negative student behaviors is the possibility that adolescents strongly dislike PE class due to social stigma. For example, overweight and obese children face strong social barriers and social isolation from their peers (Latner and Stunkard, 2003; Janssen et al., 2004). The physical demand of PE class along with the potential for increased teasing or bullying, either in the locker room or during class, may

incentivize some students to act out or skip classes altogether. This is an especially important issue if interest in school and academic performance for affected students declines in the long run.

We conclude that despite the frequent and recent recommendations for more physical activity in schools, standard PE classes are not effective in improving students well-being and may even be detrimental. Given that the TFN program was the second-largest grant program in the United States at the time of its conception, our findings have important policy implications for school spending and time allocation. In terms of cost-effectiveness, we posit that the $37 million in funding would have likely been better spent on programs such as school-based health centers if the end goal is to improve student health (Guo et al., 2010), and/or Head Start or tutoring programs that have been proven to improve student performance and close the achievement gap for low-income students (Gibbs et al., 2011). Lastly, there is scope for more work to be done on testing potential mechanisms to determine why and how physical education classes might lead to negative outcomes for middle-school students.

3. THE EFFECT OF OWN-GENDER JURIES ON CONVICTIONS

3.1 Introduction

A central right of the accused in the U.S. criminal justice system is the right to a trial before an impartial jury. This right is enshrined in the 6th amendment of the Bill of Rights to the U.S. Constitution, and was inherited from the Magna Carta, which guaranteed that no man be punished without "the lawful judgment of his peers." There are ongoing concerns, however, about the actual impartiality of juries in general, and whether jurors favor those similar to themselves in particular. These concerns have resulted in court rulings that prohibit excluding potential jurors on the basis of race, ethnicity, or sex (Batson v. Kentucky, 1986; J.E.B. v. Alabama, 1994). However, while recent research has documented bias in favor of own-race defendants (Anwar, Bayer and Hjalmarsson, 2012), there is little evidence on whether modern juries favor own-gender defendants. The purpose of this paper is to test whether own-gender juries affect criminal conviction rates and sentencing outcomes.

The primary difficulty in doing so is that seated juries are the outcome of a nonrandom jury selection process over which prosecutors, defense attorneys, and jurors have significant influence. As a result, it is difficult to distinguish the effect of own-gender juries from confounding factors, such as defense attorney quality, that lead some cases to have more jurors of the same gender as the defendant. To overcome this selection problem, we use the randomization of the initial juror pool, and the random ordering of potential jurors within that pool, to predict the proportion of female jurors seated on the jury. This enables us to use only the variation in jury gender due to the fact women are (randomly) assigned to some jury pools more than others, and women are (randomly) assigned lower numbers in the ordering of some jury pools than in others. Because the seated jury consists of the first six or twelve ordered jurors who are not excluded by either a challenge for cause or a peremptory challenge (i.e., a challenge for which no reason must be given), this variation is orthogonal to other determinants of trial outcomes. We use this quasi-random variation

in jury gender to identify own-gender effects by differencing out the impact of defendant and jury gender, similar to studies on racial bias (e.g., Price and Wolfers, 2010; Shayo and Zussman, 2011; West, 2017).

To implement this research design, we use a new data set on juror characteristics and conviction and sentencing outcomes for Palm Beach and Hillsborough counties, which are the third and fourth most populous counties in Florida. These data include all felony and misdemeanor trials over a two year period, and contain detailed information on defendant characteristics as well as case characteristics measured at both the charge and trial levels. Importantly, the data also include demographic information on potential jurors and the randomly-assigned ordering of each potential juror within the jury pool. Using this ordering and the empirical probabilities that jurors assigned a given number are seated on the jury, we predict the expected proportion of women on each jury, thereby isolating the as-good-as-random variation in the gender composition of seated juries. Importantly, we show that the predicted proportion of women on the jury is strongly predictive of the gender composition of the seated jury. We also show that this variation is uncorrelated with defendant and case characteristics, and with expected conviction rates of male and female defendants as predicted using exogenous characteristics.

Results provide strong evidence of own-gender juries on conviction rates for drug offenses. Estimates indicate that a one standard deviation increase in own-gender jurors ($\sim$10 percentage points) results in a 18 percentage point reduction in conviction rates on drug charges. Importantly, this effect is significant at the five percent level even after performing the multiple inference adjustment proposed by Anderson (2008). We also show that this change in jury gender composition leads to a 13 percentage point reduction in the likelihood of being sentenced to jail. In contrast, we find no evidence of effects for driving, property, or violent crime offenses. We hypothesize that the large effects for drug offenses are consistent with a model in which jurors are more likely to exhibit bias in cases where they have significant disagreements with U.S. law. However, we emphasize that we cannot rule out other explanations for the heterogeneity in own-gender effects. In addition, we present suggestive evidence that effects are driven largely by cases in which the jury reaches a

verdict, as opposed to cases in which a plea deal is reached prior to the trial.

To our knowledge, this is the first paper to use random variation in own-gender juries to examine effects on convictions in modern criminal courts. In doing so, the paper contributes to two literatures. The first is the broad literature examining gender bias in education, labor, housing, and product markets.[1] In addition, this paper complements a smaller body of research examining the impact of judge and jury characteristics on criminal trial outcomes. It is most similar to Anwar, Bayer and Hjalmarsson (2012) and Flanagan (2018), who show that having own-race jurors affects felony conviction rates. It is also related to Anwar, Bayer and Hjalmarsson (forthcoming), who show that while the introduction of women on English juries in 1919 had no effect on overall conviction rates, it resulted in additional convictions for sex offenses and for violent crime cases with female versus male victims.[2] This paper differs from Anwar, Bayer and Hjalmarsson (2012) and Flanagan (2018) in that we focus on jury gender, rather than jury race. In addition, it differs from Anwar, Bayer and Hjalmarsson (Forthcoming) in that we focus on the effect of jury gender in a modern context in which effects might well be significantly different than in 1919. Finally, our study differs from all three of these papers in that we observe potential juror order, rather than only the overall proportion of jurors by race or gender. This is critical when looking at the impact of jury gender, as the law of large numbers ensures that even moderately sized jury panels will have little variation in average jury gender.

In assessing the role of jurors in affecting male versus female sentencing outcomes, this paper also complements a growing literature that documents and explains gender differences in sentencing (Bindler and Hjalmarsson, 2017; Butcher, Park and Piehl, 2017). More generally, this study relates to a broader literature on the impact of judge gender (Johnson, 2014; Knepper, 2018;

[1] For example, see Abrevaya and Hamermesh (2012); Ayres and Siegelman (1995); Bagues and Esteve-Volart (2010); Bagues, Sylos-Labini and Zinovyeva (2017); Breda and Ly (2015); Dahl and Moretti (2008); De Paola and Scoppa (2015); Goldin and Rouse (2000); Lavy (2008); Neumark, Bank and Van Nort (1996); Moss-Racusin, Dovidio, Brescoll, Graham and Handelsman (2012)

[2] While we focus on the effect of own-gender juries in this paper, we also examine the effect of jury gender composition on overall conviction rates. Results are shown in Appendix Table A1, in which we regress an indicator for conviction on our measure of expected proportion women on the jury. Overall, we find no evidence that additional female jurors are more or less likely to convict overall.

Schanzenbach, 2005; Steffensmeier and Hebert, 1999) and other judge and jury characteristics.[3] Finally, in documenting how defendants who draw opposite-gender juries are more likely to be convicted and sentenced, this paper also complements recent papers documenting unfairness in conviction and sentencing based on other factors (Eren and Mocan, 2018; Philippe and Ouss, 2017).

The results of this study have important implications. First, they suggest that even in settings where participants are actively reminded of the importance and necessity of being fair and impartial, sizeable gender biases can still occur. In addition, we note that there is strong evidence that higher conviction and incarceration rates lead to increased recidivism and worsened labor market outcomes (Aizer and Doyle Jr, 2015; Mueller-Smith, Forthcoming). As a result, our findings suggest that drawing an opposite-gender jury can impose significant long-run costs on defendants.

3.2 Background and Data

3.2.1 The assignment of potential jurors to the jury pool and the voir dire process

As described above, a critical feature of our research design is the random assignment of residents to panels of potential jurors, and the random ordering of residents within each panel. In the Florida counties we study, county court offices randomly mail jury summons to residents who have a driver's license or identification card. Potential jurors arrive at the courthouse on the assigned day and enter their information into a computer system. Each potential juror is then randomly assigned to a case. In addition, within each case each potential juror is assigned a number.

The potential jurors for a given case are then escorted to the courtroom for the voir dire, or jury questioning, process. As described by U.S. Supreme Court Justice Rehnquist, "Voir dire examination serves to protect [the right to an impartial jury] by exposing possible biases, both known and unknown, on the part of the jurors. Demonstrated bias in the responses to questions on voir dire may result in a juror's being excused for cause; hints of bias not sufficient to warrant challenge for cause may assist parties in exercising their peremptory challenges" (McDonough

[3] Examples include Anwar, Bayer and Hjalmarsson (2014; 2015); Mitchell, Haw, Pfeifer and Meissner (2005); Cohen and Yang (forthcoming); Depew, Eren and Mocan (2017); and George (2001).

Power Equipment, Inc. v. Greenwood, 1984, page 464). Prosecutors and defense attorneys are allowed unlimited challenges for cause, though meeting the requirements for removing a potential juror is difficult, and such requests are not always granted by the judge. In Hillsborough and Palm Beach Counties, each side is typically allowed up to three peremptory challenges to remove jurors they believe unlikely to be favorable toward their side of the case. The final jury thus consists of the first six or twelve jurors not struck by either side, beginning with the potential juror assigned number one. Any remaining potential jurors are then excused or returned to jury services to be reassigned.

3.2.2 Data

We obtained detailed administrative data for all misdemeanor and felony cases that were assigned potential juror pools in preparation for trial in Palm Beach and Hillsborough Counties from 2014 to 2016.[4] These are the third and fourth largest counties in Florida, respectively, each with a 2016 population of over 1.3 million people. Importantly, these data include comprehensive information on the voir dire process along with case attributes. Specifically, we observe the pool of jurors randomly assigned to each case including name, seat number, and outcome of the selection process.

Data from Hillsborough County also include the gender of potential jurors, as well as date of birth, race, and address. For Palm Beach County, we infer gender on the basis of the first name. We do so using an online application programming interface called genderize.io. The application predicts gender based on first name using a large dataset comprised of user profiles from several major social networks. Using this approach, we are able to predict probabilistic genders for 92% of potential jurors. For the names that we do not predict, we assign 0.5 to the female gender indicator variable under the assumption that the missing name is equally likely to be male or female. To verify the accuracy of this approach to inferring gender, we compare predicted gender to actual observed gender in Hillsborough County, and find that we accurately predict gender 94.38% of the

[4] There are 32 cases in Palm Beach County and 1 case in Hillsborough County where there should be a jury panel but the information was not in the case file. Only two of these cases involve drug related charges.

time. We then combine potential juror order and the gender of each potential juror to predict the number of women we would expect to serve, on average, for each trial.

From these data we are able to compute empirical probabilities of being seated on the jury for each spot in the order in the jury panel.[5] To do so, we let the probability vary by size of the potential jury pool and the number of jurors being selected. Standard juries in Florida consist of six jurors, though the judge may decide to seat 12 jurors in some cases. Importantly, this decision is made prior to the assignment of the jury pool, and thus should not affect the internal validity of our approach. These probabilities are shown in Figure B.1, where panel a shows the probability of being seated on the jury for six-person juries, and panel b shows the same for twelve-person juries. For example, for six-person jury trials with a panel size of 20 or less, the probability of being selected for the jury is around 40 percent for the first 10 or so potential jurors, and then declines to around 20 percent for the 20th-ordered potential juror. By comparison, for 12-person juries selected from panels of 50 to 100 potential jurors, the probability of being seated ranges from 25 to 30 percent for the first 40 jurors to close to zero for the potential juror assigned last (e.g., 100th) in the jury pool.

To predict the number of women that will be seated on the jury, we interact the estimated probabilities shown in Figure B.1 with a gender indicator variable equal to one for females.[6] Summing this over the pool of potential jurors gives the expected number of females seated. Since trials in our data consist of both six and 12 person juries, we divide by the jury size to get the expected proportion of females. This enables us to make meaningful comparisons across jury panel sizes. In Section 3.1 we demonstrate that the predicted proportion of women on the jury is highly correlated with the actual proportion of women on the jury. In addition, in Section 3.2 we show that the expected proportion of women on the jury is uncorrelated with case and charge characteristics,

[5] In some cases, a second panel of potential jurors was used. Our understanding is this sometimes occurred because the first panel did not result in enough seated jurors, and sometimes because the judge chose not use the first panel at all for some reason. However, we still observe the first (and subsequent) juror panels in those cases, and we order the jurors accordingly. For example, if each of the first two panels had 50 potential jurors, we assign number 51 to the the first ordered juror in the second panel, and number 100 to the last juror in that second panel. We do so even if no jurors from the first panel were seated on the jury.

[6] The probability of a potential juror being female is used for panels in Palm Beach County.

which is consistent with random assignment to panels and random ordering within panels.

For each case in our data, we observe the charges brought against the defendant and the outcome of each charge including verdict and sentencing. Our primary outcome of interest is an indicator for whether the defendant is convicted of the charge. Importantly, our data include guilty and innocent verdicts issued for all cases for which a jury panel was assigned in preparation for trial. For example, we observe guilty pleas that arise after the jury pool was assigned as well as verdicts found by the jury. This precludes the possibility of selection bias, since some cases settle after the prosecutor or defense attorney observes the composition of the potential jury pool or the actual seated jury. In addition, we note that for some charges in Florida, a verdict can be given in which adjudication is withheld. In that case the defendant is assigned a term of probation, and upon successful completion of that term is spared a conviction on his or her record. This is the outcome in only 3.56 percent of all charges in our sample, and only 4.05 percent of drug charges. For the main analysis we treat this outcome as guilty, though in Table B.7 we show that estimates are similar if we instead classify it as not guilty. Our second outcome of interest is whether and for how long a defendant is sentenced to be incarcerated upon the conclusion of the trial. We define this outcome at the trial level, rather than charge level, since the sentences of individual charges are often served concurrently. In each case we observe the defendant's gender and race along with additional case characteristics including the severity of charges and the judge assigned to the case.

Finally, we note that because the purpose of this paper is to examine the effect of own-gender juries, we exclude cases linked to charges in which fewer than 10 percent of defendants are female. Consequently, we only consider cases that involve a drug, driving, property, or violent crime. In addition, we limit violent crimes to domestic crimes, assaults, and robberies. This is due to the low number of female defendants in other violent crime categories, such as sexual assault and murder, which gives us very little variation in defendant gender.

Summary statistics are shown in Table B.1, where Panel A shows characteristics at the trial level, and Panel B shows characteristics at the charge level. We have a total of 1,542 cases/defendants, representing 3,055 separate charges. Sixty-seven percent of defendants are convicted of at least

one charge, while men are convicted at somewhat higher rates than women (67 versus 63 percent). Across all cases, on average defendants are sentenced to 1,673 days in jail, though men are sentenced for significantly longer than women (1,931 versus 268 days).[7] Sixteen percent of our defendants are female, 48 percent are white, and the average age is 37.

3.3 Methods

In order to identify the effects of own-gender juries, we use a generalized difference-in-differences approach. Specifically, we estimate the following linear probability model:

$$\begin{aligned} Convict_{ct} =& \beta_1 DefFemale_t + \beta_2 E(PropFemale)_t + \beta_3 DefFemaleXE(PropFemale)_t \\ &+ X_t + County_t + CountyXCrime_{ct} + \epsilon_{ct} \end{aligned} \tag{3.1}$$

where the outcome of interest $Convict_{ct}$ is a binary variable equal to one if the defendant is convicted guilty of charge c in trial t. $DefFemale_t$ an indicator variable equal to 1 if the defendant in trial t is female, controls for differences in conviction based on defendant gender. Similarly, $E(PropFemale)_t$, the expected proportion of females seated on the jury for trial t, accounts for differences in the decision to convict due to the gender of jurors. The coefficient of interest, β_3, measures the effect of own-gender juries on the outcome. X_t is the set of control variables at the trial level including defendant's age and race, the total number of charges against the defendant, if the case involves a violent charge, the predicted age of the jury pool, and judge gender. All specifications include county fixed effects along with county-by-crime fixed effects when considering more than one crime category. Observations are weighted by the inverse of the total number of charges in a trial.

Robust standard errors are clustered at the defendant level to allow for errors to be correlated across charges and trials for a given defendant. In addition, because we also test for the presence of own-gender juries by crime severity (felony vs. misdemeanor), and by crime type (drug, driving, property, and violent), we also report False Discovery Rate (FDR) adjusted Q-values. These are

[7] We assign a sentence of zero days to those defendants who are sentenced to time served.

computed using the method proposed by Anderson (2008), and adjust for the fact that we examine effects on conviction for six different categories of crime.[8] These are interpreted similarly to p-values from a two-tailed test, and explicitly adjust for the increased likelihood of estimating extreme coefficients when making multiple comparisons.

In addition, we also test whether the effect of own-gender juries on conviction translates to differences in sentencing, which is decided by a judge rather than the jury. For that reason, we focus primarily on the category of charges for which we find an effect on conviction. Due to the discrete nature of prison sentences, the presence of many zero observations, and the wide dispersion of sentence lengths, we estimate the effect of own-gender juries on the distribution of sentences using binary indicators. This is done at the trial level as the sentences for individual charges are often served concurrently. Formally we estimate the following ordinary least squares regression for each binary sentence length:

$$\begin{aligned} AtleastXDays_t =& \beta_1 DefFemale_t + \beta_2 E(PropFemale)_t + \beta_3 DefFemaleXE(PropFemale)_t \\ &+ X_t + County_t + CountyXCrime_t + \epsilon_t \end{aligned} \tag{3.2}$$

where $AtleastXDays_t$ is a binary indicator for X days sentenced in trial t with X starting with at least 1 day and increasing by 6 month increments to 10 years. The covariates are defined as in the previous equation where β_3 is interpreted as the degree of own gender juries. We allow for correlation in errors among trials with the same defendant by clustering at the defendant level.

The intuition of this generalized difference-in-differences approach is to compare the difference in how male and female defendants are judged by less-female juries to the difference in how male and female defendants are judged by more-female juries. This approach allows more-female juries to convict at different rates than more-male juries, so long as this difference is constant across male and female defendants. Equivalently, we allow male defendants to be "more guilty" than female

[8] While we also test for sentencing effects, the focus of the paper and therefore the multiple inference adjustment is on convictions. This is because in these counties, conviction is the only outcome over which juries have direct control. In these counties, sentencing decisions are made by judges based on those conviction outcomes.

defendants, though we require that this difference in underlying guilt be similar for more-male and more-female juries.

The identifying assumption of this approach is that while male defendants may have different underlying likelihood of conviction than female defendants, in the absence of a treatment effect the difference in their conviction rates should be the same for more-male juries as for more-female juries. This assumption could be violated in a couple of different ways. The first is if our measure of jury gender is correlated with other factors that affect conviction rates. For example, if skilled defense attorneys are able to strike opposite-gender jurors at higher-than-average rates, then we might observe lower conviction rates when there are more same-sex jurors and falsely attribute it to own-gender juries. To overcome this problem, we construct a measure of expected jury gender composition that is based on the random assignment of individuals to jury pools and the random ordering of individuals within the jury pool. We show that this measure of jury gender is both strongly correlated with the composition of the seated jury, and is orthogonal to other observed determinants of conviction rates such as defendant and case characteristics. We also show that the difference in the guilt propensity of male and female defendants, as predicted using all exogenous characteristics, does not vary with the gender composition of the jury.

The second way in which the identification assumption can fail is if female jurors tend to be more likely to convict defendants of certain crimes (or when certain other crimes are also being charged), and if those crimes are disproportionately committed by certain genders. For example, if women are more likely to convict on a theft charge when a violent crime was also committed at the same time, and if male defendants are more likely than female defendants to be charged with both theft and violent crime, this approach could overstate the effect of own-gender juries. Similarly, if women are more likely than men to convict blacks, and if there is a higher proportion of black male defendants than black female defendants, then our estimated could be biased. To address this possibility, we show the robustness of our estimates to the inclusion of controls that interact the (expected) gender composition of the jury with various case characteristics, such as race and whether the defendant is also being charged with a violent crime. In addition, we include

controls that interact the gender composition of the jury with other defendant characteristics, such as race. If the inclusion of these interactions were to result in a decline in our estimate of interest, it suggests that at least some of the effect is due not to own-gender bias, but to differential treatment of some other defendant characteristic correlated with defendant gender.

3.4 Results

3.4.1 Correlation between expected jury gender and actual jury gender

We begin by demonstrating that our measure of jury gender, which is the expected proportion of women on the jury based on the random potential juror assignments and orderings, is predictive of actual jury composition. Note that in contrast to the main analysis, this exercise can only be performed for those cases in which a jury was seated for the trial. The underlying data are shown in Figure B.2, which graphs the actual proportion of women seated on the jury against the expected proportion of women seated. It shows strong positive correlations for both 6-person juries and 12-person juries. In both cases the slope is close to one, suggesting that our (exogenous) measure of jury gender composition is strongly correlated with observed jury gender composition.

Regression results are shown in Table B.2. Specifically, we estimate an equation of the same form as equation (1) above in that we regress the actual proportion of females on the predicted proportion of females, along with county-by-crime fixed effects. Results are consistent with Figure B.2 in showing strong correlations between actual and expected gender composition. Column 1 shows a correlation of 0.949, significant at the 1 percent level, for all case types. The remaining columns show that this correlation remains strong for felonies, misdemeanors, and cases that include infractions related to drugs, driving, property crime, and violent crime. Correlations range from 0.860 for driving cases to 1.042 for misdemeanor cases. All estimates are statistically significant at the 1 percent level. As a result, it is clear that the combination of more women being assigned to a jury pool and being assigned earlier in the ordering leads to large subsequent differences in the actual gender composition of the seated jury.

3.4.2 Exogeneity tests of the measure of expected jury gender composition

The validity of our empirical approach depends in large part on the assumption that predicted jury gender composition is uncorrelated with confounding factors. While we expect this assumption to hold based on our understanding of how potential jurors are assigned to and ordered within jury pools, we can also provide some empirical evidence. To do so, we regress exogenous defendant and case characteristics on the expected proportion of jurors who are female. These characteristics include jury panel size as well as defendant gender, race, age, the number of offenses, and whether the defendant is being charged with a felony, drug, driving, property, or violent crime. In addition, we also test whether average juror age (available only for Hillsborough County) or judge gender is correlated with our measure of the expected proportion of women on the jury.

Results are shown in Table B.3, with estimates at the trial level shown in Panel A, and at the charge level in Panel B. Overall, there is little evidence that these exogenous characteristics are correlated with our measure of expected jury gender composition. Of the 24 estimates shown, two are significant at the 10 percent level, and one is significant at the five percent level, which is consistent with random chance. This contrasts with results from the same exercise using actual proportion of women on the seated jury, rather than our measure of expected jury gender composition. In that exercise, the results of which are shown in Appendix Table B.9, nine of the 24 estimates are significant at the 10 percent level, and three are significant at the five percent level.[9] This reflects the fact that the actual proportion of women seated for the jury is the outcome of the non-random jury selection process.

In addition, we also provide another test. The intuition of the test is to use all of the exogenous case and defendant characteristics shown in Table B.3, along with county-by-crime fixed effects, to predict conviction rate for each charge for each individual. This predicted conviction rate is thus a linear combination of all observable characteristics about that case and individual, where the weights are optimally chosen to best predict the likelihood of being convicted on that charge.

[9] In cases where no jury is seated, we assign actual proportion female to be the expected proportion female. If we instead limit the sample to those trials in which jurors were seated, six estimates are significant at the 10 percent level with five estimates significant at the 5 percent level.

We graph these predicted conviction rates for male and female defendants against our measure of expected jury gender composition. Our identifying assumption requires that the difference in the underlying propensity for guilt of male and female defendants be orthogonal to jury gender.

Results for all charges are shown in Figure B.3a. The symbols represent local averages for charges against male and female defendants, and are grouped into 10 equal-sized bins. In addition, we fit separate lines to the underlying data for male and female defendants. Figure B.3a shows that while male defendants are predicted to be found guilty more often than female defendants, this difference is constant across jury gender. This suggests that there is little reason, based on observable case and defendant characteristics, to expect a nonzero difference-in-differences estimate in the absence of an effect of own-gender juries.

Results in Figure B.3b show predicted conviction rates for drug charges, where we later show large effects of own-gender juries. Results are similar to Figure B.3a in that while male defendants are predicted to have higher conviction rates than female defendants, this difference does not vary with expected jury gender. This is consistent with the identifying assumption, and suggests that any nonzero difference-in-difference estimate of the effect of jury gender will be due to the effect of jury gender, rather than some confounding factor.

3.4.3 Effect of own-gender juries on conviction rates

Next, we turn to estimating the effect of jury gender on convictions. Before presenting formal estimates, we first show the raw data. Figure B.4 graphs the conviction rates of male and female defendants against the expected proportion of females on the jury. Results for all charges are shown in Figure B.4a. It shows that the conviction rates of male defendants are relatively flat as the expected proportion of female jurors increases. By comparison, the conviction rates of females seem to decline somewhat as the expected proportion of female jurors increases, though the difference in slopes is relatively subtle.

Conviction rates for drug offenses are shown in Figure B.4b. Conviction rates for male defendants appear to increase somewhat as the expected proportion of female jurors increases. In contrast, conviction rates of female defendants decline sharply as the expected proportion of fe-

male jurors increases. The locally averaged conviction rates for female defendants facing juries with an expected proportion of females less than 0.5 range between 60 and 100 percent. By comparison, locally averaged conviction rates for juries expected to be more than half female range from 20 to 50 percent. In short, female defendants are much less likely to be convicted of a drug charge as the jury is more female, while if anything men are more likely to be convicted as the jury is more female.

Estimation results are shown in Table B.4. All specifications control for the expected proportion of female jurors as well as an indicator for whether the defendant is female. In addition, all specifications control for county-by-crime fixed effects. Column 1 shows the estimate of own-gender juries for all crimes. The coefficient is -0.247 and is not statistically significant. The magnitude of the coefficient implies that a 10 percentage point change in the expected gender of the jury is associated with a 2.47 percentage point reduction in the conviction rate.

Column 2 additionally controls for other defendant and case characteristics such as the defendant's age and race, judge gender, the number of charges in the case, and whether the defendant was also charged with a violent crime such as assault. Consistent with the identifying assumption, the coefficient changes little to -0.256 and remains insignificant.

As discussed earlier, a major threat to identification is the possibility that more male or more female juries are responding not to defendant gender, but to a feature of the case or defendant that is systematically correlated with defendant gender. For example, if women convict at higher rates for all charges when the defendant is also charged with a violent crime, and if male defendants are more likely to be charged with violent crimes along with other crimes, then we can estimate a nonzero own-gender effect even if women apply this standard equally across all defendants. In order to address this concern, in the third column we examine the robustness to our estimate to the inclusion of controls that interact case characteristics with defendant gender and the expected proportion of female jurors. Specifically, we include interactions of the proportion of female jurors with defendant race, age, judge gender, number of charges in the case, whether the individual is being charged with a violent crime, and whether the defendant is being charged with a felony. This

allows for the possibility that jurors are responding differentially to defendant characteristics that may be correlated with defendant gender.

Results from a specification that includes these pairwise interactions are shown in column 3 of Table B.4. As shown there, the coefficient of interest becomes somewhat larger at -0.329, though is still statistically insignificant.

Columns 4-6 of Table B.4 show results for felonies. Estimates range from -0.321 to -0.468, though none are statistically significant at conventional levels. Similarly, results in columns 7 âĂŞ 9 show results for misdemeanor charges. Again, all estimates are negative ranging from -0.469 to -0.485 and none are statistically significant. Importantly, due to the fact that we report results for several different subcategories of crime, we also report False Discovery Rate (FDR) adjusted Q-values for each estimate in Table B.4. These are computed using the method proposed by Anderson (2008), and adjust for the fact we examine a total of six subcategories of crime (felony, misdemeanor, drug, driving, property, and violent). The adjusted Q-values, which are interpreted similarly to two-sided p-values, range from 0.531 to 0.657 for the estimates in columns 4-9.

Next, we examine effects by category of the criminal charge. Specifically, we examine effects on conviction for driving, property, violent, and drug crime charges. Results are shown in Table B.5. The format is similar to Table B.4 in that the first column for each category includes only county fixed effects, the second column adds controls for defendant and case characteristics, and the third column adds controls for interactions between jury gender and defendant and case characteristics.

Results in columns 1-9 suggest there is little evidence that own-gender juries affect convictions for driving, property, or violent crimes. In contrast, results in columns 10-12 indicate there is strong evidence of own-gender juries on conviction for drug charges. The estimate of -2.205 in column 10 suggests that a 10 percentage point change in the expected own-gender composition of the jury results in a 22 percentage point reduction in the conviction rate of defendants. Adding controls changes the estimate only slightly to -2.192, and further adding interaction controls results in an estimate of -1.815. All estimates are statistically significant at the one percent level. More

importantly, FDR-adjusted Q-values are 0.002, 0.002, and 0.078, respectively. This indicates that even after accounting for the multiple statistical tests across the six major categories of crime charges in Tables B.4 and B.5, the coefficients in columns 10-12 of Table B.5 are sufficiently extreme as to be unlikely to arise due to chance.

To put these estimates in perspective, we note that Anwar, Bayer and Hjalmarsson (2012) estimate that the impact of having one black potential juror in the jury pool (and thus likely less than a 10 percentage point increase in the expected proportion of jurors that are black) results in a 16 percentage point reduction in the conviction rates for black defendants.

3.4.4 Effects of own-gender juries on sentencing decisions

Next, we turn to the question of whether own-gender juries affect sentencing. While one may expect increased convictions to result in additional incarceration, we note that this link is *a priori* ambiguous for two reasons. The first is that the additional convictions may be for charges that do not result in incarceration. In addition, while juries make conviction decisions, in these counties judges decide sentencing. On the one hand, if judges treat all convictions similarly, we would expect to observe own-gender effects on sentencing for drug cases. On the other hand, if judges exercise discretion in sentencing based on either the facts of the case or even on the gender composition of the jury, we may not see evidence of own-gender effects in sentencing outcomes.

Results are shown in Figure B.5, with panels a and b showing results for all cases and drug cases, respectively. Each panel shows estimates of the effect of own-gender juries in which the outcome of interest is whether the defendant was sentenced for at least one day, at least six months, at least one year, at least 18 months, etc., up to at least 10 years. Results for all cases shown in Figure B.5a indicate that while there is some evidence that own-gender juries resulted in reduced sentences — especially on the left-hand side of the distribution — none of the estimates are statistically significant.

Results in Figure B.5b indicate there is a statistically significant decline in the likelihood of receiving a sentence of at least one day. Estimates for the effect on longer sentences are positive but not statistically significant. This suggests that juries are less likely to convict those own-gender

defendants who might otherwise be convicted and sentenced to relatively short sentences.

These results are shown more formally in Table 6, which shows estimates of the effect of own-gender juries on the probability of being sentenced to at least some jail time. Consistent with Figure B.5a, estimates in columns 1-3 for all charges are negative but not statistically significant. In contrast, estimates for cases that include at least one drug charge shown in columns 4-6 range from -1.264 to -1.453, and are all statistically significant at conventional levels. These estimates imply that a 10 percentage point change in the expected gender composition of the jury results in a 13 to 15 percentage point change in the likelihood of being sentenced to jail or prison.

These findings have several important implications. First, they suggest that own-gender juries do lead to differences in sentencing outcomes, even when sentencing decisions are made by judges. This mean that judges are either unwilling or unable to exercise discretion in an effort to offset the effect of jury gender composition on conviction decisions. In addition, the effects on sentencing imply that not only does drawing an opposite-gender jury lead to a criminal record, but it also leads to increased incarceration. Existing research on the effect of conviction and incarceration on recidivism and employment suggests that this results in significant long-term harm to defendants on drug charges (Aizer and Doyle Jr, 2015; Mueller-Smith, Forthcoming).

3.5 Robustness

As discussed earlier, a major threat to identification of own-gender jury effects is the possibility that jurors of a given gender are responding not to the defendant's gender, but to some other defendant or case characteristic correlated with defendant gender. We test for this by including interactions of jury gender with the number of charges in the case, whether there was a charge for a violent crime in the case, judge gender, and defendant race and age. Results in column 12 of Table B.5 indicate our estimates are robust to the inclusion of these interactions, which provides evidence that the effects are due to the interaction of jury and defendant gender and not something else. However, one may also be concerned that jurors of different gender could respond differently to the type of drug charge in the case, which could be correlated with defendant gender. To test for this, we additionally include interactions of expected jury gender with indicators for marijuana

possession, possession of other drugs, and possession of drug paraphernalia, where drug trafficking is the excluded group. Results are shown in column 2 of Table B.7, where column 1 replicates our main estimate for drug charges of -1.815 from column 12 of Table B.5. Results in column 2 show that including these interactions increases the magnitude of the estimate to -2.092. This provides further evidence that the effects shown are due to the interaction of defendant and jury gender, rather than the interaction of jury gender with some other characteristic correlated with defendant gender.

In addition, we also test the robustness of our estimates to different specifications as well as to alternative ways of constructing our predicted jury gender measure. In column 3 of Table B.7 we estimate the effect controlling for predicted juror age, which we only observe in Hillsborough County. The estimate is similar at -2.296, and is significant at the five percent level. Column 4 shows the estimate from our main specification when we classify outcomes in which adjudication was withheld as not guilty rather than guilty, which occurs in 4.05 percent of the drug charges. The magnitude of the estimate is reduced slightly to -1.659, but is still statistically significant at the 5 percent level.

In columns 5-7 of Table B.7, we estimate the own-gender jury effect when we classify the gender of potential jurors differently. Specifically, we classify jurors for whom we could not identify gender using genderize.io as either all female (column 5) or all male (column 6), respectively, rather than as having an equal likelihood of being male as female. Estimates are similar in magnitude and significance at -1.614 and -1.844, respectively. In addition, in column 7 we classify the gender of jurors based on the names and genders recorded in Florida by the Social Security Administration. The resulting estimate is -1.749, which is similar to the baseline estimate of -1.815.

Finally, in columns 8-10 of Table B.7 we show that our estimate of own-gender juries is robust to alternative methods of predicting jury gender. In column 8 we estimate the effect when we do not smooth the probability of being seated on the jury for a given jury and panel size using a local linear estimation with epanechnikov kernel, as we did for our main results. Instead, we use the raw probability that a juror assigned that number in a panel in a given range was seated on the jury. The

estimate is -1.681 and is significant at the five percent level. The estimated effect is also similar if we use probit instead of local linear estimation, as shown by the estimate of -1.887 in column 9. The same is true when we use a local linear smoother but do not condition on jury panel size (-1.698), as shown in column 10.

In summary, we find no evidence that our estimated effect of own-gender juries on convictions in drug cases is due to male or female jurors responding differentially to a characteristic correlated with defendant gender, rather than defendant gender itself. In addition, we find that this own-gender effect is robust to alternative ways of defining the outcome and predicting jury gender.

3.6 Discussion and Interpretation

There are several potential mechanisms through which own-gender juries could have such large effects on conviction and sentencing outcomes. The first is that seated jurors may exhibit own-gender bias when making conviction decisions on drug charges. Given that we do not observe true guilt, it is difficult for us to assess which jurors — male or female — are biased, and in what direction. But under this interpretation, the results would be due to male and/or female jurors being either too lenient to own-gender defendants, being too tough (i.e., wrongfully convicting) on opposite-gender defendants, or both.

Relatedly, effects could be due to the expectation of juror bias in criminal drug trials. For example, a defendant may be more likely to accept an otherwise unappealing plea deal if the expected jury composition is largely opposite-gender. It is also possible that prosecutors or defendants falsely believe jurors will engage in gender bias during the trial, resulting in a change in plea deal behavior prior to the start of the trial.

Finally, an increase in the number of opposite-gender jurors could lead the defense to use their peremptory challenges on opposite-gender potential jurors. This would mean the attorney would have fewer peremptory challenges to use on other unfavorable jurors, thereby weakening the defendant's chances at acquittal. However, we note that doing so would violate the legal standard set by Batson v. Kentucky (1986) and J.E.B. v. Alabama (1994). In addition, the fact that predicted jury gender is so highly correlated with actual jury gender provides empirical evidence that the

attorneys are unable to significantly offset random changes in expected jury gender.

Data limitations make it difficult for us to distinguish between these potential mechanisms with any certainty. However, to shed some light on this question, we estimate effects separately for cases that did and did not get to trial.[10] Results are shown in Appendix Table A3, which shows that both sets of point estimates are statistically significant at the one percent level. However, the magnitude of the effect for charges decided by jury is twice as large as the effect when the case was decided prior to the conclusion of the trial. We interpret this as suggestive evidence that effects are largely driven by changes during or after the trial, such as gender bias by juries. We note, however, that selection into whether a case goes to trial after the jury panel is assigned makes it difficult to interpret these differences with certainty.

A second question regarding the interpretation of this studyâĂŹs findings relates to the strength of the effects for drug charges compared to driving, property, and violent crime. Unfortunately, our data are not well-suited for explaining this difference across crime types. We speculate it is because even though Americans are supportive of existing and even stronger penalties for DUIs, violent crime, and property crime, Americans are critical of the prosecution of drug crimes. For example, recent surveys indicate that 40 percent of Americans believe the prison sentences for non-violent drug crimes are too harsh, and 64 percent support the full legalization of marijuana (YouGov/Huffington Post, 2015; Gallup News Service, 2017). Two-thirds of American adults believe the government should focus more on treatment for illegal users, compared to only 26 percent who believe more focus should be on prosecuting illegal users (Pew Research Center, 2014). A nontrivial proportion of Americans even disagree with the prosecution of "harder" drug crimes; 16 percent favor decriminalization of cocaine possession, and 9 percent favor legalization (Morning Consult, 2016). This shift in attitudes on drug laws is also reflected in recent state policy changes regarding drug possession.[11] These views are particularly relevant given the drug charges

[10] For this analysis we exclude the 67 cases representing 150 charges where the records did not indicate whether the case was decided by trial or prior to the start of the trial.

[11] The National Conference of State Legislatures (NCSL) reports that from 2011 to 2016, at least nine states have lowered some drug possession crimes from felonies to misdemeanors, and another nine have reduced mandatory sentences for some drug offenders (National Conference of State Legislatures , NCSL). In addition, as of 2018 over 20 states have decriminalized certain marijuana possession offenses (National Organization for the Reform of

in our sample, over 58 percent of which are for possession of drugs or drug paraphernalia without intent to distribute.

In contrast, there is little to no public support for weakening the enforcement of non-drug laws, and significant support for even strengthening enforcement. While surveys of Americans' perceptions of non-drug offense prosecution are less common, what evidence there is contrasts sharply with views on drug crime enforcement. For example, only 11 and 1 percent of adults believe that the sentences typically given for non-violent property crimes and violent crimes, respectively, are too harsh (Huffington Post/YouGov, 2013). As a result, we interpret this study's findings as most consistent with a model in which jurors fairly enforce the laws with which they mostly agree, but disproportionately favor own-group defendants when deciding whether to enforce laws with which they might not agree. That is, while a juror may be willing and able to convict out-group defendants who break a law with which the juror disagrees, she is perhaps less willing to convict in-group defendants of the same crime. We emphasize, however, that there could be other explanations for the difference in results for across crime types.

3.7 Conclusion

In this study, we test for the effect of own-gender juries on conviction and sentencing outcomes. To overcome potential bias due to nonrandom jury selection, we exploit the fact that potential jurors are randomly assigned to jury pools for each case, and are randomly ordered within each jury pool. This enables us to predict the gender composition of each jury for each case set to go to trial, thereby isolating the as-good-as-random variation in jury gender. We combine this variation with variation in defendant gender to estimate the effect of own-gender juries.

Results provide strong evidence that own-gender juries result in lower conviction rates for drug offenses. We estimate that a ten percentage point change in the expected own-gender composition of the jury results in a 18 percentage point decline in conviction rates on drug charges. A similar change in jury gender results in a 13 percentage point reduction in the likelihood of being sentenced

Marijuana Laws , NORML). While Florida is not among the states making these changes, jurors there are likely experiencing similar shifts in their views about drug laws.

to at least some jail time. These are large effects, though we note this is consistent with prior research on the effect of juror race (Anwar, Bayer and Hjalmarsson, 2012).[12]

We hypothesize that the reason we see such strong own-gender effects for drug charges but not others is because many Americans disapprove of the prosecution of drug crimes. We emphasize, however, that we cannot rule out other interpretations. Similarly, while we show evidence that effects are largest for cases that go to trial, it is difficult for us to determine which part of our effect is due to gender bias by jurors when deciding to convict, and what is due to changes in the offering or acceptance of plea deals based on perceptions of jury bias.

Our results are important for the debate over the use of peremptory challenges in selecting a jury. By documenting the significant harm that can arise to defendants who draw opposite-gender juries, we highlight the potential benefits to the prosecution of removing same-gender individuals from the jury pool. Similarly, defendants in drug cases stand to benefit greatly if their attorneys are able to successfully remove opposite-gender jurors from the jury pool. As a result, our results provide support for recent court rulings that disallow prosecutors or defense attorneys to strike potential jurors from the jury pool on the basis of gender.

In addition, our results add evidence to a growing literature documenting own-gender bias in decision-making. Our findings suggest that such bias can arise even in settings where the objective of impartiality is heavily emphasized and protected. Specifically, throughout the juror selection process the necessity of being impartial and fair is actively pressed on potential jurors. In addition, the process explicitly allows for both sides to remove potential jurors from the jury if they are shown or believed to be unfair. We find that even in this process, the similarity in gender of the jury to the defendant has a significant effect on conviction and sentencing outcomes.

[12] They find that one black individual in the jury pool — and thus in expectation much less than one black juror on the seated jury âĂŞ results in a 16 percentage point change in conviction rates.

4. THE IMPACT OF ECONOMIC OPPORTUNITY ON CRIMINAL BEHAVIOR: EVIDENCE FROM THE FRACKING BOOM

4.1 Introduction

Since Becker (1968), crime has been viewed as the outcome of rational individuals weighing costs and benefits of legal and illegal forms of employment. Thus, if individuals face improved labor markets, the returns to legal activity increase and individuals should substitute away from illegal activities. Yet, local economic booms are often associated with increases in crime (Grinols and Mustard, 2006; Freedman and Owens, 2016; **?**). Several theories can rationalize this phenomenon including increases in criminal opportunities, access to disposable income for activities that complement crime, and population changes. However, the extent to which each of these theories explains this puzzle is unclear, especially since changes in crime are typically observed at an aggregate level.

The purpose of this paper is to address this puzzle by estimating the effect of local economic opportunity on the criminal behavior of residents who already lived in the area prior to the economic boom. By focusing on the criminal behavior of existing residents, I distinguish the effect of economic opportunity from the effect of the compositional changes in the population caused by in-migration during the boom. This is important, as people tend to leave as labor market conditions worsen and migrate to areas during economic expansions. I do so by using the recent boom in hydraulic fracturing in North Dakota as a large, exogenous shock to an individual's relative returns to legal versus illegal behavior. This approach, combined with the focus on the behavior of residents already living there prior to the start of hydraulic fracturing, enables me to identify the effect of economic opportunity on individual criminal behavior.

I identify effects using a difference-in-differences framework comparing counties located in the shale play, to counties not located in the shale play over time. Importantly, I measure the impact on residents, separating out migration effects, using information on local residents prior

to the economic shock. The sharp increase in hydraulic fracturing activity in the United States is an ideal economic shock for several reasons. First, areas were affected based on the formation of the shale play beneath the Earth's surface. Second, the shock was largely unforeseen, as fracking suddenly became a viable method due to a combination of technological innovations (Wang and Krupnick, 2013; Crooks, 2015). Together, these support the assumption that fracking affected local labor markets for reasons unrelated to prior local conditions and household behaviors, overcoming common critiques of the difference-in-differences design.[1] Third, hydraulic fracturing was large enough to affect the entire local economy in many areas. Finally, the shock affected predominately low-skill jobs, a population of policy interest.

Studying the effect at the individual level requires detailed data on hydraulic fracturing activities, criminal behavior, and local residents in North Dakota.[2] I identify residents in each county from printed directories in the early 2000s before the in-migration associated with production activities. As an important measure of criminal behavior, I obtained detailed administrative data on the universe of criminal cases filed in the state from 2000 to 2017. I also observe which residents signed a lease and received royalty payments during this period. This enables me to not only identify the effect of improved labor market opportunities, but also isolate differential effects of those receiving large, non-labor income shocks and those that do not. Matching these datasets makes it possible to study the effect of local economic shocks on the criminal behavior of local residents. This is an important advantage given the large migration effects that have been documented in response to economic conditions in general, and to fracking in particular (Wilson, 2017).

Results indicate that the start of the economic expansion — defined as the period when companies began leasing mineral rights and investing in the area — led to a statistically significant 0.44 percentage point (22 percent) reduction in criminal behavior by local residents. Effects are largest for drug-related crimes, though I also see some less precisely estimated declines in other

[1] For example, see Besley and Case (2000) for discussion about policy endogeneity in difference-in-differences frameworks.

[2] North Dakota is well suited for this analysis as it was the third-slowest-growing state in 2000, and increased its real gross domestic product 115% by the end of the fracking boom in 2016 (U.S. Bureau of Economic Analysis, 2018). Also, it is the second largest crude oil producing state in the United States.

crimes. The effect is smaller once production began, with a 0.27 percentage point decrease in the likelihood of committing a crime and not consistently statistically significant. This suggests that changes during the production period, such as increased income or changes in peer composition, offset some of the effect of improved job opportunities.

In addition, I exploit variation in mineral rights ownership and royalty income to assess the extent to which effects are driven by labor market opportunities versus non-labor income shocks. Results indicate that the reduction in crime seems to be driven by non-leaseholders. This is consistent with those not receiving income through alternative means being more responsive to increased job opportunities. However, I note that the effect sizes are not statistically distinguishable.

To my knowledge, this is the first paper to identify effects of economic shocks on individuals' criminal behavior separate from the effect of migration. In doing so, it contributes to two bodies of literature. First, it contributes to the literature showing how aggregate crime changes in response to plausibly exogenous shocks to economic conditions (e.g., Dix-Carneiro, Soares and Ulyssea, Forthcoming; Axbard, 2016; Grinols and Mustard, 2006; Gould, Weinberg and Mustard, 2002; Raphael and Winter-Ebmer, 2001; Evans and Topoleski, 2002; Montolio, 2018; Grieco, 2017). These studies generally show aggregate crime is inversely related with economic conditions, with some exceptions.

In particular, this paper complements a subset of this literature that has documented the role of criminal opportunity and income inequality in explaining the observed increases in aggregate crime that arise during economic expansions (e.g., Mejia and Restrepo, 2016; Cook, 1986). For example, Freedman and Owens (2016) study the effect of BRAC funding in San Antonio on crime using individual-level data. They find an increase in property-related crime in neighborhoods with a high composition of construction workers, those most likely to benefit from the economic shock, and that crime is more likely to be committed by individuals with a prior criminal record, and thus unable to be employed by the project. In a similar way, this paper documents that once one accounts for population changes that accompany economic expansions, one observes the expected relationship between improved job opportunity and individual crime. Together, the findings of

those papers and this paper suggest that both criminal opportunity as well as shifts in population can explain the puzzling finding that aggregate crime shifts during economic expansions.

Second, this study contributes to the growing literature on the effects of fracking, which has transformed many regions in the United States. Specifically, crime has generally been shown to increase in areas with fracking activities (Alexander and Smith, 2017; Andrews and Deza, 2018; Komarek, 2017; Bartik, Currie, Greenstone and Knittel, 2016).[3] However, the increase could be driven by changes in the population of workers moving to the area or an individual's response to the changing economic conditions. I measure a similar increase in aggregate cases filed in fracking counties, but find that local residents in the county are actually less likely to commit crime when exposed to relatively stronger labor market conditions. This is consistent with predictions of the economic theory of crime when the returns to legal employment increase, and indicates that fracking has reduced individuals' propensity to commit crime.[4]

Finally, while the primary purpose of this study is to examine the role of economic expansions on the criminal behavior of local residents, this study's findings on aggregate crime also speak to the literature on (im)migration and crime. Immigration to the United States and Western Europe typically increases in response to improved relative economic opportunity in those countries. Many worry that the immigration to high-income countries could increase crime rates, though some recent empirical evidence suggests this fear may be misplaced (Bell, Fasani and Machin, 2013; Chalfin, 2015; Spenkuch, 2013; Miles and Cox, 2014; Butcher and Piehl, 2007).[5] Results on aggregate crime presented here indicate that the migration of mostly young, American men does lead to increased crime overall. Thus, changing the composition of a local population can be an

[3] Alternatively, Feyrer, Mansur and Sacerdote (2017) do not find statistically significant evidence of an increase in crime across all counties with fracking.

[4] This is also consistent with empirical evidence documenting a similar inverse relationship between recidivism and economic conditions (e.g., Agan and Makowsky, 2018; Yang, 2017; Galbiati, Ouss and Philippe, 2018; Schnepel, 2017), as well as increased lifetime criminal behavior for cohorts graduating high school in harsher economic conditions (Bell, Bindler and Machin, Forthcoming)

[5] While the overall evidence on this question is mixed, Bell, Fasani and Machin (2013) finds no effect on violent crime and mixed effects on property crime, Chalfin (2015) shows an increase in aggravated assaults, but decreases in other crimes, and Spenkuch (2013) reports small increases in crime, particularly financial crime. Relatedly, Miles and Cox (2014) finds no effect of a deportation policy on local crime. Moreover, Butcher and Piehl (2007) shows that immigrants typically have lower crime rates than do native-born residents potentially due to a combination of heavy screening of would-be migrants, and self-selection of those migrants.

important driver of criminal activity, although the effects may depend heavily on who the migrants are. Since young men are a particularly crime-prone population, economic booms that attract this group may be more likely to lead to higher crime rates.

4.2 Background

Advances in hydraulic fracturing contributed greatly to the recent oil boom in the United States. From 2000 to 2015, oil produced from fractured wells increased from 2% to over 50% of domestic production, increasing total oil production faster than at any other point in time (EIA 2018a). It suddenly became more profitable due to a breakthrough in directional drilling, hydraulic fracturing technologies, and seismic imaging (Wang and Krupnick, 2013; Crooks, 2015). Hydraulic fracturing involves injecting fluids at a high pressure into a shale play in order to crack the rock formation and extract tight oil and shale gas.[6] This process allowed mineral resources to be extracted from shale plays that were previously not economically viable.

Notably, the Bakken formation in North Dakota, smaller only than the Permian and Eagle Ford formation in Texas in crude oil production, is one such formation. Figure C.1 shows the 17 counties that produce oil and gas in North Dakota, classified by production levels as either a core (major) or balance (minor) county.[7] Four counties make up the major fracking counties producing 80% of North Dakota's oil from 2000-2017, with the remaining 13 producing 20%.

Companies leased the mineral rights required for production from individuals or agencies in exchange for a portion of total revenue. Figure C.2a plots the number of leases signed by households in North Dakota each year from 2000 to 2017. It is clear that lease signing spiked in 2004 signaling when companies first began investing in hydraulic fracturing in North Dakota. Similarly, Figure C.2b graphs total oil production in North Dakota showing that production lagged leasing by a few years, starting to increase in 2008. From 2008 to 2017, North Dakota produced oil valued at

[6] The Energy Information Administration defines a shale play as a "fine-grained sedimentary rock that forms when silt and clay-size mineral particles are compacted, and it is easily broken into thin, parallel layers. Black shale contains organic material that can generate oil and natural gas, which is trapped within the rock's pores" (2018). I focus on oil production as North Dakota's production is typically only 10-20% gas, with the rest being oil.

[7] Figure is reprinted from North Dakota's Oil and Gas by Report North Dakota Labor Market Information (2018).

an estimated $2,904,191 million dollars.[8]

Perhaps unsurprisingly, the presence of hydraulic fracturing activities has had a substantial impact on local labor markets. Feyrer, Mansur and Sacerdote (2017) estimate that every one million dollars of new production generates 0.78 jobs and $66,000 in wages in counties with a shale play across the United States.[9] Similarly, in response to the stronger labor markets, Wilson (2017) estimates that the in-migration of workers increased the baseline population in fracking counties by 12% on average in North Dakota. Additionally, individuals who also owned mineral rights received 10-20% percent of production revenues through royalty payments. As I show in the next section, I estimate that the average leaseholder earned a royalty of $11,000 per month, which is a substantial non-labor income shock.

Figure C.3 shows that prior to the fracking boom, counties in North Dakota are relatively similar in terms of per capita income, total jobs, population, and total officers. The leasing period, when residents first knew of the fracking boom, was characterized by slight increases in per capita income, total jobs, and population (2004 to 2008). Oil production began ramping up in 2008 and is the more labor-intensive period. This is when companies began offering high paying jobs and moving in a large number of workers, often into camps due to housing shortages. It is also the period when the majority of households that had signed a lease received royalty payments and increases in overall crime were reported. This is reflected in the data, as Figure C.3 shows fracking counties experienced large increases in income, jobs, population, and police officers during the post-2008 production period. While the economic opportunities continued through this period, counties changed in several other ways as well. As a result, in my analysis I will estimate the effect of expected economic opportunity that occurs after signing but before drilling, as well as the effect of drilling. I expect the former will pick up the effect of job opportunities both expected and realized, while the latter will measure the effect of job opportunities along with large population

[8] This estimate is calculated based on total monthly oil production in North Dakota and the monthly North Dakota oil first purchase price.

[9] Other papers estimating increases in wages and employment from fracking activities include Bartik, Currie, Greenstone and Knittel (2016); Allcott and Keniston (2017); Fetzer (2014); Maniloff and Mastromonaco (2014); Weber (2014) and Gittings and Roach (2018) to name a few.

and income changes.

Economic theory predicts that the labor market changes from fracking activities may affect crime in several ways. First, the additional jobs and higher wages should induce individuals to substitute away from illegal activities now that the returns to legal activities are higher. Alternatively, the large cash transfers — via royalty payments — to some households may lead to more crime through increased income inequality and opportunity of crime. Additionally, the increased income through either royalties or higher wages could affect crime by easing financial constraints or providing more disposable income to consume goods that may complement crime (e.g. alcohol). Finally, the large migration effects observed in the production period are likely to affect crime both through population increases and compositional changes.

There are three main advantages of studying the effect of positive economic shocks on crime in this context. First, the sudden increase in hydraulic fracturing activities creates plausibly exogenous variation in exposure to improved labor market conditions. Second, I am able to distinguish the effect on crime by the existing population from aggregate effects which include individual changes in behavior as well as compositional changes. Specifically, I am able to focus my analysis on households already living in the area using directory files in each county to identify residents. Finally, I can study how these residents respond to changes in economic opportunity as well as the economic opportunity plus the subsequent influx of people and income by examining both the earlier leasing period and the more labor-intensive production period.

4.3 Data

For this analysis, it is necessary to identify residents in years prior to the fracking boom to account for migration. To do this, I collect a list of all rural residents for each county in North Dakota prior to 2008 from the Great Plains Directory Service.[10] Households listed in these directories represent roughly 20% of all households in North Dakota during this time. The directory

[10] All counties are included except Cass, Grand Forks, Pembina, and Traill, which are not covered by the Great Plains Directory Service during this time.

information includes the name, address, and city of all rural residents.[11] In total, there are 31,169 households defined by resident last name, street number, city, and zip code. I consider this to be the universe of households for which I match to lease and crime data using a Levenshtein index.[12]

One potential concern with identifying residents is that some people may have moved into fracking counties prior to the large in-migration associated with the production period. For example, strategic households may move in advance to have first access to housing or jobs. However, to be recorded in the resident directories, any movers would have to move into the rural areas. If this were the case, we would expect to see an increase in property sales prior to the production period. I show in Figure C.8 that property sales in fracking counties remain similar to sales in non-fracking counties throughout the leasing period. Thus, the residents in directory files are likely all long-time residents of the county.

All leases spanning from 2000 to 2017 in North Dakota are collected from Drilling Info, a private company designed to aid companies participating in all steps of mineral production. Data include name and address of the grantor, company listed as grantee, number of acres leased, royalty rate, and date of record. Production data at the county- and well-level are collected from the North Dakota Department of Mineral Resources. I use these datasets to approximate the amount of monthly oil production from a given well that is attributed to an individual leaseholder. This amount is dollarized using the North Dakota Crude Oil First Purchase Price to estimate the amount leaseholders receive in the form of royalty payments.[13]

The State of North Dakota Judicial Branch provided restricted administrative data on all crim-

[11] Notably, residents living within city limits are not included in the directories and thus are not considered in this analysis.

[12] I allow matches with a string distance of 2 or less. In practice, this means two strings are matched across datasets if there are only two changes that need to be made to the concatenated string of last name, street number,city, and zip code in order for them to be exact matches. In Table C.6, I show that main results are robust to this index.

[13] Each well in North Dakota is assigned a spacing unit which defines the area of land surrounding the well with rights to production. These boundaries are determined in court hearings at the request of the proposed well operator and based on recommendation of geologists. By matching leaseholders to spacing units, I define the proportional interest in monthly production for each leaseholder based on acres leased. The dollar value is calculated using the monthly North Dakota Crude Oil First Purchase Price which I subtract $10/barrel to account for post production costs, namely transportation. I deduct 10% for severance tax, as North Dakota collects 5% for gross production in lieu of property tax on mineral rights and 5% for oil extraction. Leaseholders then get a fraction of depending on their negotiated royalty rate, typically 12-18%.

inal cases filed in North Dakota from 2000 to 2017. Importantly, data include identifying information including the name, date of birth, and address of individuals charged with a crime. This allows me to link to residential files and identify crime committed by local residents. I also observe the file date, specific charges filed, disposition of each charge, sentence received, and county of filing for every case.

There are two main advantages to using cases filed as a measure of criminal behavior. First, it is considerably more serious than 911 calls or arrests, as an individual has officially been charged with a crime. This is reflected by the fact that only 61% of all arrest charges in North Dakota were filed by the prosecutor's office over the last five years (North Dakota Attorney General Office 2018). As a result, charges filed are arguably a less noisy measure of criminality than the other possible alternatives. Additionally, the State of North Dakota specifically advises employers not to ask about prior arrests as "an arrest does not mean that someone actually committed a crime" (North Dakota Department of Labor and Human Rights 2018). Second, since cases filed are recorded in an administrative database, they do not suffer from voluntary reporting practices or a lack of coverage, particularly in areas that are sparsely populated. Additionally, these data report information on all charges, including offenses which are often not tracked in other commonly used datasets such as drug charges or driving while under the influence.

Summary statistics are shown in Table C.1. Close to 20% of households are ever charged with a crime from 2000 to 2017 (Table C.1, Panel A). The types of charges filed for this population, namely rural residents, are summarized in Panel C. The majority of crimes are misdemeanors ($\sim$90%), with driving-, drug-, and property-related charges making up roughly 38%, 17%, and 14.5% of all charges, respectively. Smaller crime categories representing less than 10% of all charges, such as assault (4%), are grouped together in other charges. Of these households, roughly 20% sign a lease and may receive royalty payments during this period (Table C.1). Close to 40% of leaseholders in my sample actually received payments during this period, with the average leaseholder receiving $11,000 per month. These royalty payments can be thought of as an additional treatment over the local economic shock, as some residents in fracking counties receive large,

additional lump sums of money while others do not.

4.4 Methodology

4.4.1 Main analysis

The unexpected rise in fracturing activities coupled with spatial variation in the shale play provide a plausibly exogenous shock to local economic conditions. Using a generalized difference in differences framework, I compare the criminal behavior of residents in counties within the shale play to residents in counties outside the shale play before and after the fracking boom.[14] Given the timing of fracking activities and subsequent changes in affected counties, I consider the effects separately in each period: leasing (2004 to 2008) and production (2008 to 2017). Formally, I estimate the effects of local economic shocks on criminal behavior with the following linear probability model:

$$\begin{aligned} CriminalBehavior_{ht} &= \alpha_h + \gamma_t \\ &+ \theta_1 FrackingCountyXPostLease + \theta_2 FrackingCountyXPostProduction_{ht} + \epsilon_{ht} \end{aligned} \tag{4.1}$$

where criminal behavior is a binary indicator for whether a case was filed for household h in year t.[15] In some specifications, criminal behavior is separated into the four largest crime categories: property, driving, drug and other. Household fixed effects, α_h, account for any static differences in the propensity to commit crime across households. Year fixed effects, λ_t, control for factors that affect criminal behavior for all households in a given year, such as the Great Recession. $FrackingCountyXPostLease_{ht}$ and $FrackingCountyXPostProduction_{ht}$ are indicator variables equal to 1 for households in fracking counties during the leasing period and during the production period, respectively. Here, θ_1 and θ_2 are the coefficients of interest measuring the difference in criminal behavior of residents in fracking counties relative to residents in non-fracking

[14] I also report aggregate county-level estimates of equation 1 in appendix Figure C.10 for comparison.

[15] Since the dependent variable is binary, I additionally show results using a logistic regression in Table C.7. I also show results for the intensive margin using both the number of individual cases filed and the total number of charges in a given year using the Inverse Hyperbolic Sine (IHS) transformation and Poisson models.

counties in each of the treatment periods.

The assumption behind this approach is that residents' criminal behavior in fracking counties would have changed similarly over time with residents' criminal behavior in non-fracking counties, absent hydraulic fracturing activities. I check this assumption in several ways. First, I provide visual evidence that treated and control counties are tracking prior to any treatment. Relatedly, I formally test for pre-divergence using the above regression model with an indicator for the treated group one year before treatment. Additionally, I allow counties to trend differently over time by including county-specific linear time trends. I also include interactions between pre-treatment controls and year effects. In doing this, I allow for counties with different levels of observable characteristics, such as per capita income, to respond differentially to year-to-year shocks.

In all models, robust standard errors are clustered at the county level, allowing errors to be correlated within a county over time. I also report permutation-based inference for the primary specification when considering all crime, similar in spirit to Abadie, Diamond and Hainmueller (2010) for inference when using the synthetic control method. For this, I randomly assign treatment to 17 counties and compare the estimated coefficient to 1000 placebo estimates to compute p-values. In addition, I report False Discovery Rate (FDR) Adjusted Q-values when estimating effects separately by crime type (property, driving, drug, and other) following Anderson (2008). FDR Adjusted Q-values correct for the increased likelihood of rejecting the null hypothesis when making multiple comparisons, and are interpreted similar to p-values.

Given that some counties have larger shocks than others, detected effects could be driven solely by counties with more extreme local shocks. However, it is beneficial to know if smaller economic shocks also affect criminal behavior. Therefore, I also consider heterogeneous effects by the amount of fracking activity experienced by a county. Specifically, I estimate the treatment effect for the four major oil and gas producing counties as defined by the Labor Market Information Center, namely Dunn, McKenzie, Mountrail, and Williams, separate from the effect in the thirteen minor fracking counties.

4.4.2 Effects by leaseholder status

Finally, I examine the potentially differential effects of fracking on leaseholders and non-leaseholders. As previously discussed, some households receive large sums of money in the form of royalty payments while others do not. This creates the potential for increased crime due to changes in both income inequality and criminal opportunities. I consider leaseholders and non-leaseholders within fracking counties as separate treated groups, comparing each of them to residents in non-fracking counties. To the extent that signing or not signing a lease and receiving royalty payments is also a form of treatment, this strategy separates the effect on the two groups living in fracking areas. Formally, I estimate the following regression model:

$$\begin{aligned} CriminalBehavior_{ht} &= \alpha_h + \lambda_t \\ &+ \beta_1 LeaseHolderXPostLease_{ht} + \beta_2 LeaseHolderXPostProduction_{ht} \\ &+ \phi_1 NonLeaseholderXPostLease_{ht} + \phi_2 NonLeaseholderXPostProduction_{ht} + \epsilon_{ht} \end{aligned} \tag{4.2}$$

where variables are defined as in equation 1. Now, β_1 and β_2 measure the change in criminal activity by leaseholders in fracking counties compared to residents in non-fracking counties during fracking activities. They capture both the effect of job opportunities and the additional income received by leaseholders in the form of royalty payments. Similarly, ϕ_1 and ϕ_2 measure changes in criminal activity by non-leaseholders in fracking counties to residents in non-fracking counties. Alternatively, they capture the effect of higher wages and job opportunities, along with any potential effect of not receiving royalty payments for non-leaseholders. As in the previous models, equation 2 is estimated using two periods: leasing starting in 2004 and production beginning in 2008. Notably, leaseholders receive a small signing bonus upfront with royalty payments closely following production, as leaseholders receive a percentage of production revenues.

4.5 Results

4.5.1 Main results

I begin by estimating the overall effect of local economic shocks on crimes committed by residents. As noted above, I consider only the population of residents prior to the fracking boom in North Dakota. In doing so, I am able to exclude all crimes committed in the county by new workers who migrated to the relatively stronger labor markets. In this way, I can distinguish the effect of the economic shock from the impact of the changing demographics on overall crime rates.

First, I graph the estimated divergence over time in crimes committed by residents in fracking and non-fracking counties, relative to the difference between the two sets of counties in 2000 and 2001. Figure C.4 plots the dynamic difference-in-differences estimates for all crimes, controlling for household and year fixed effects. Importantly, there is no evidence of divergence prior to the start of the fracking boom in 2004. This supports the identifying assumption that absent hydraulic fracturing activities, residents in fracking counties would have experienced similar changes in criminal behavior as residents not in fracking counties. Additionally, the figure indicates that the probability of being charged with a crime falls in fracking counties when leasing starts, then rises some during the production process. This suggests economic opportunity is reducing crime, but the effect seems to be offset at least somewhat by the indirect effects that accompany oil production. For example, the production period also includes interactions with new workers and increases in disposable income from royalty payments or high-paying drilling jobs. I report the average treatment effects for each period in Table C.2.

Starting with the leasing period, Column 1 indicates an initial drop of 0.44 percentage points in overall crime by residents in fracking counties relative to residents in non-fracking counties. This translates to a 22% drop in cases filed and is statistically significant at the 1% level. Moreover, the permutation-based p-value is less that 1% with 1 out of 1000 placebo estimates less than -0.0044, shown graphically in Figure C.11. In Column 2, I formally test for pre-divergence and find no evidence of it, with the coefficient on the lead indicator being close to zero, -0.0009, and

statistically insignificant. In column 3, I allow for county-specific linear trends. This allows for both observable and unobservable county characteristics to change linearly over time. If results are driven by fracking counties being on a different path than non-fracking areas, then adding a county-specific linear trend should absorb the treatment effect. However, results indicate the coefficient increases slightly to -0.50 percentage points. Finally, counties with different baseline populations, total jobs, officers, per capita income, and production may respond differentially to year-to-year shocks. For example, if fracking counties also tend to be smaller in population then detected effects could be a result of small counties differentially responding to yearly shocks. In Column 4, I allow these observable characteristics in 2000 to differentially affect criminal behavior each year. The magnitude remains stable at -0.50 percentage points. Notably, all coefficients are statistically significant at the 1% level, and the estimated effect is robust to the inclusion of various controls and a lead term.

Overall, estimates in Table C.2 are consistent with Figure 4 in showing that while there is a significant drop in criminality initially, the drop is somewhat diminished in the production period. Column 1 indicates a 0.27 percentage point reduction in cases filed for residents in fracking counties compared to residents in non-fracking counties, although not statistically significant at conventional levels. The permutation-based p-value is marginally significant at 8.8%, with 88 of the 1000 placebo estimates less than or equal to the estimated coefficient. Moving across columns 2 through 4, coefficients remain negative ranging from -0.18 to -0.43 percentage points, and only two of which are significant at the 10 percent level. It appears as though the reduction in criminal behavior from the boost in economic activity may be at least somewhat offset by additional effects on criminal behavior during the production period. This could be due to the effects of in-migration such as peer effects and increased social interaction (Glaeser, Sacerdote and Scheinkman, 1996; Ludwig and Kling, 2007; Bernasco, de Graaff, Rouwendal and Steenbeek, 2017), or to an increase in the number of bars and illegal markets.[16]

To better understand the type of crime affected by local economic shocks, I present treatment

[16] This is graphically depicted in Figure C.9 with a large increase in the average total number of liqour licenses per county in counties with major fracking activity.

effects separately for financial-related crimes (e.g. theft, criminal mischief, fraudulent checks), driving-related crimes (e.g. DUIs, reckless driving), drug-related crimes (e.g. possession), and other crimes (e.g. assault, resisting arrest, criminal conspiracy). The dynamic difference-in-differences estimates, controlling for household and year fixed effects, are plotted for each crime type in Figure C.5. Notably, the figures show that residents in fracking and non-fracking counties do not diverge prior to the fracking boom in these types of crime. However, residents exposed to fracking activities change their criminal behavior relative to residents in non-fracking counties in response to the economic shock. Results show relatively large reductions in driving, drug, and other offenses after the leasing period. However, this reduction is diminished once production starts.

I follow the same format as Table C.2 reporting average treatment effects for each period in Table C.3, with panels for each crime type and reported FDR Q-values for statistical inference. Panel A indicates a -0.06 to -0.16 percentage point decrease in property cases filed during the leasing period, and a -0.11 to -0.20 percentage point decrease during the production period. Similarly, estimates are negative for driving-related cases during the leasing (-0.21 to -0.22 percentage points) and production period (-0.02 to -0.09 percentage points). Panel C shows a decrease in drug cases filed of -0.20 to -0.28 percentage points during the leasing period, and a reduction of 0.04 to -0.15 percentage points during the production period. Finally, all other crimes have a similar negative effect during the leasing period ranging from -0.15 to -0.25, with a smaller effect once production began ranging from -0.04 to -0.22 percentage points. All coefficients are fairly robust to the inclusion of controls and a lead term.

Because I consider four types of crime, I also report statistical significance of these estimates using the Adjusted False Discovery Rate Q-values proposed in Anderson (2008). These values correct for the increased chance of rejecting the null hypothesis when making multiple comparisons for two treatments across four groups (eight categories). The negative effects on driving and property cases are generally not statistically significant once corrected for multiple comparisons. However, the effect on drug cases filed during the leasing period is sufficiently large across all

specifications in Column 1 through 4 as to not have occurred by chance with Q-values of 0.076, 0.049, 0.062, and 0.052, respectively, and no statistical effect during the production period. The effect on all other cases is less robust with two of the four FDR Q-values less than 0.10 during the leasing period, again with no estimated effect once production began.

For comparison, I also report the effect of hydraulic fracturing activities on aggregate changes in cases filed per 1000 persons. Appendix Figure C.10 plots the dynamic coefficients from the county level model of equation 1, with county and year fixed effects, for all cases and by case type. Again, counties do not diverge prior to fracking activities. However, estimates indicate increases in total cases filed, as well as drug, driving, assault, and all other cases during the fracking periods, specifically during production, which is consistent with prior literature.

Additionally, I test whether the migrants entering the fracking counties were committing crimes at higher rates than the native population. This enables me to speak directly to a question of interest in the immigration literature of whether those moving into an area are more criminogenic in general. I measure the propensity to commit crime for a subset of those moving into the county. Specifically, I calculate the crime rate using the number of cases filed with an out-of-state address over the number of migrant tax exemptions filed in the county. Similarly, I do the same for all crime committed by those with an address in North Dakota and the number of non-migrant tax exemptions in the county, fixing the total as of 2000. I find that the crime rate from 2004-2015 is higher for those moving into the county at 17%, as measured by crime committed by out-of-state individuals, compared to a rate of 7% for in-state individuals.[17]

Taken together, findings provide strong evidence of a reduction in residents' criminal behavior during the leasing period; these effects seem to be partially reduced by other effects during the production period. While all crime types are negative, results are primarily driven by drug-related

[17] Notably, this can be thought of as a conservative estimate. First, the crime rate for people moving into the county only considers crime from out-of-state individuals even though there is some in-migration to fracking counties from other areas in North Dakota. This also means that any additional crimes committed by those that move into the county from within the state are being considered as crimes committed by non-migrants for this exercise. Second, migrant and non-migrant tax exemptions are based on whether there is a change in filing county and state. The denominator for out of state is the total of all inflows from 2000 through 2015 to be conservative. Similarly, I fix the total number of non-migrants in each county at the total in 2000 for each year as migrants that move into the area will be counted as non-migrants in their second year residing in the county.

crimes. This contrasts the county-level results, suggesting that compositional changes play an important role in the criminal response to economic conditions. Put differently, this suggests that the aggregate increases seen are due largely to additional crimes committed by those who move into the area. In contrast, the effect of the economic opportunity itself seems to have a negative effect on overall crime.

4.5.2 Results by intensity

Results thus far have treated all counties on the shale play as receiving the same economic shock. However, some counties experience much larger economic shocks than others, particularly the four major oil and gas producing counties. Specifically, the oil production in each of these four counties was greater than the amount produced in the other 13 counties combined over this time period. To estimate the differential effect by treatment intensity, I report estimates from equation 1 separately for major and minor fracking counties in Table C.4. Following the same format as Table 2, I first discuss estimates for minor and major fracking counties during the leasing period, 2004 to 2008. Estimates in Column 1 indicate a 0.39 percentage point decrease in cases filed by residents in counties with minor fracking activity and a 0.55 percentage point decrease in the major fracking counties, significant at the 5% and 10% level respectively. This represents a 19.5% reduction in cases filed in counties with minor fracking activity and a 27.5% reduction in the major fracking counties. The estimated effect is stable to the inclusion of a lead indicator, county specific trends, and allowing for time-shocks that vary with levels of pre-period observables. Estimates in Columns 2 to 4 range from a 0.39 to 0.44 percentage point decline in minor fracking counties and 0.60 to 0.75 in major fracking counties. All estimates are significant at conventional levels.

During the production period, estimates for minor fracking counties are similar in magnitude ranging from a 0.19 to 0.40 percentage point reduction in cases filed, although marginally significant. Estimates for major fracking counties are smaller in magnitude during the production period relative to the leasing period (-0.07 to -0.62 percentage points), and not consistently significant at conventional levels.

As expected, the effect is larger in magnitude for the major fracking counties than in minor

fracking counties initially, although coefficients are not statistically different. Importantly, this demonstrates that the effect is not driven solely by the four large fracking counties, as counties experiencing more modest economic shocks also see a significant reduction in crime. Additionally, the effect seems to fade more dramatically in the major fracking counties which also experience larger population and income changes during the production period. This is consistent with the interpretation that it is the other consequences of the in-migration, such as peer effects, and income that lead to a diminished reduction in crime for residents.

4.5.3 Results by lease-holder status

In addition to the local economic shock, some residents in fracking counties also receive a large positive income shock in the form of oil royalties during the production period. Recall that the average household that signs a lease receives over $10,000 dollars per month from royalty payments. These payments may affect the decision to commit crime both for the leaseholder and the non-leaseholder as payments increase disposable income for illegal activities by leaseholders while increasing the income inequality and criminal opportunities for non-leaseholders. In Table C.5, I estimate the extent to which the fracking activities may differentially affect residents using equation 2.

Estimates for lease-holders are all negative during leasing (-0.21 to -0.28 percentage points) and production (-0.08 to -0.27 percentage points), although none are significant at conventional levels. Estimates for non-lease-holders range from -0.66 to -0.70 percentage points during the leasing period and are all significant at the 1% level. During the production period, estimates range from -0.34 to -0.59 for non-lease holders, with three of the four estimates significant at the 5% level. While these estimates are not statistically different from each other, it is clear that the overall reductions in crime shown in Table 5 are primarily driven by reductions in crime by those who do not receive royalty payments. This suggests that it is the increase in job opportunities that reduces crime, rather than income per se. Moreover, the effect of job opportunities seems to be stronger than the effect of increased criminal opportunities.

4.6 Discussion

In summary, I find that crime decreases during the leasing period in response to improved job opportunities, and that the effect is partially reduced once drilling activities escalate throughout the production period. Importantly, the effect is not driven by the four largest oil producing counties, and the fact that the effect shrinks more in these counties suggest that other factors related to production contribute to offsetting the effect of improved labor market conditions. Additionally, I find that effects are strongest for non-leaseholders and persist into the production period. This is consistent with those not receiving alternative income streams being most sensitive to the job opportunities.

One concern in interpreting the results described above is that the differences over time may be due to changes in the number of police. Becker (1968) and others highlight that the probability of detection factors into an individual's decision to commit crime, which is also echoed in the lab (Harbaugh, Mocan and Visser, 2013). Moreover, empirical evidence has shown that crime decreases in response to increased police presence (di Tella, 2004; Machin and Marie, 2011). To test for changes in the police force, I estimate the main model at the county level with total police officers as the outcome of interest. Figure 6a, indicates that change in the amount of police officers was negligible during the leasing period. As a result, changes in police are unlikely to be driving the significant reduction in crime observed during the leasing period. However, changes in police are potentially part of the treatment during the production period, although this is difficult to disentangle from other factors that changed during that period. Similarly, reductions in police resources from population increases may lead to fewer reported cases filed (Vollaard and Hamed, 2012). Figure 6b shows little evidence of changes in the population from 2004 to 2008, with large increases during the more labor-intensive production period. Again, population changes are less of a concern during the leasing period, but are likely to be a part of the treatment effects after 2008 as previously discussed.

Relatedly, a concern may be that people identified as residents may have moved out of the county or, more importantly, the State of North Dakota during the fracking periods. This could be

an issue if changes in crime are simply from not observing the criminal behavior of an individual that moved out of the state. Anecdotally, it seems improbable that residents would disproportionately move out of fracking counties as economic conditions improved. I empirically check for evidence of out-migration using the number of tax exemptions that move out of a county each year. I find no evidence of differential out-migration during the initial leasing period and only signs of out-migration toward the end of the production period when those that had moved into the county begin leaving as shown in Figure C.7

While I am not able to directly test for the mechanism underlying the decrease in crime from improved economic opportunities, I suggest two potential pathways. First, it is possible that decreases in crime are the result of an incapacitation effect, as individuals become occupied with legal work and thus have less time for criminal activities. This is similar to school having an incapacitation effect on juvenile crime (Jacob and Lefgren, 2003). A second explanation is that residents may no longer feel the need to engage in activities related to crime, such as drug use, given their improved economic outlook. This is consistent with work by Case and Deaton (2015; 2017) and Autor, Dorn and Gordon (Forthcoming), who document an increasing number of deaths from drugs, alcohol and suicide associated with deteriorating economic conditions. This is also consistent with Becker (1968) which predicts individuals are less likely to engage in criminal activity if they have more to lose if apprehended. As a result, a more positive outlook on economic conditions, whether expected or realized, may also lower crime.

4.7 Conclusion

This paper studies the effect of local economic shocks on individualsâĂŹ decisions to commit crime. Specifically, I exploit the recent boom in hydraulic fracturing activities as a plausibly exogenous shock to local economic conditions. Using detailed administrative data on all criminal cases filed in North Dakota from 2000 to 2017, I estimate the effect of increased job opportunities on criminal behavior. An important strength of this study is that by focusing the analysis on all rural residents already living in the area prior to fracking, I can distinguish the effect of improved economic opportunity from the effect of population inflows on aggregate crime.

Results indicate that, consistent with the existing literature, aggregate crime increased in fracking counties relative to non-fracking counties. This was particularly true during the more labor-intensive fracking activities. However, local residents engage in less criminal activity at the start of the boom with a smaller effect in later years. Effects are largest and most robust for drug offenses, and are shown across all counties with fracking activity. Additionally, I show that effects are most pronounced for residents that do not also receive royalty payments. Taken together, results suggest that residents reduce their criminal activity in response to improved job opportunities, but that other changes from local economic shocks, such as peer composition, seems to reduce this effect. This is consistent with economic opportunities reducing crime and highlights the role of compositional changes on the aggregate effects on crime.

5. SUMMARY AND CONCLUSIONS

In this dissertation, I have presented three essays which provide empirical evidence to inform social and policy motivated questions covering a variety of topics in labor economics. The first essay in Section 2 documents the effects of a physical education grant program mandating daily PE attendance which was targeted toward low-income middle schools. Causal effects are estimated by exploiting the eligibility threshold for the grant using a regression discontinuity design. Results indicate that the program was not effective at improving student fitness or achievement. However, there is some evidence that the program may have been detrimental to student behavior with increases in disciplinary infractions and decreases in attendance.

The role of own-gender bias in criminal jury trials is examined in Section 3. Given that jurors are selected, the gender composition of the potential jury pool, which is both randomly assigned and ordered, is used to isolate the effect of gender and overcome any potential selection bias. Intuitively, trials that have a larger share of women randomly assigned earlier in the potential jury pool are quasi-randomly more likely to have women on the final seated jury. This allows for the effect of having more women or men on a jury to be estimated with a focus on the differential effects when the defendant is male or female. Findings show that a standard deviation increase in the share of own-gender jurors results in a significant decrease in the likelihood of conviction and being sentenced to any time in jail for drug-related charges only.

The final chapter, presented in Section 4, studies how individuals respond to improved economic opportunity with respect to their criminal behavior. To do so, I exploit the recent fracking boom in North Dakota as a source of exogenous variation in local economic conditions. I identify residents living in the area prior to changes in conditions and study their behavioral response to the improved labor markets relative to residents living in areas not exposed to hydraulic fracturing activities. In doing so, I am able to separate changes in aggregate crime which includes individual changes in behavior as well as changes in population. While fracking activities are associated with overall increases in local crime, I find a significant decrease in crime committed by those who

were already living in the area that now face better economic conditions. Moreover, these effects are most pronounced for drug-related charges and for non-leaseholders suggesting individuals are responding more to the increased job opportunities than income per se.

Together these studies help inform policies related to health, education, the criminal justice system, discrimination, stimulus programs, and hydraulic fracturing. As data becomes increasingly available, it is possible to shape more useful policies using rigorous economic research which I hope I am able to contribute to with these and future studies.

REFERENCES

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of CaliforniaâĂŹs tobacco control program," *Journal of the American statistical Association*, 2010, *105* (490), 493–505.

Abrevaya, Jason and Daniel Hamermesh, "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?," *Review of Economics and Statistics*, 2012, *94* (1), 202–207.

Agan, Amanda and Michael Makowsky, "The Minimum Wage, EITC, and Criminal Recidivism," 2018. Working Paper.

Aizer, Anna and Joseph Doyle Jr, "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges," *Quarterly Journal of Economics*, 2015, *130* (2), 759–803.

Alexander, James and Brock Smith, "There will be blood: Crime rates in shale-rich U.S. counties," *Journal of Environmental Economics and Management*, 2017, *84*, 125–152.

Allcott, Hunt and Daniel Keniston, "Dutch disease or agglomeration? The local economic effects of natural resource booms in modern America," *The Review of Economic Studies*, 2017, *85* (2), 695–731.

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

Andrews, Rodney and Monica Deza, "Local natural resources and crime: Evidence from oil price fluctuations in Texas," *Journal of Economic Behavior & Organization*, 2018, *151*, 123–142.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson, "The Impact of Jury Race in Criminal Trials," *Quarterly Journal of Economics*, 2012, *127* (2), 1017–1055.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson, "The Role of Age in Jury Selection and Trial Outcomes," *Journal of Law and Economics*, 2014, *57* (4), 1001–1030.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson, "Politics in the Courtroom: Political Ideology and Jury Decision Making," *Journal of the European Economic Association*, 2015.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson, "A Jury of Her Peers: The Impact of the First Female Jurors on Criminal Convictions," *Economic Journal*, Forthcoming.

Autor, David, David Dorn, and Hanson Gordon, "When Work Disappers: Manufacturing Decline and the Falling Marriage Market Value of Young Men," *American Economic Review: Insights*, Forthcoming.

Axbard, Sebastion, "Income Opportunities and Sea Piracy in Indonesia: Evidence from Satellite Date," *American Economic Journal: Applied Economics*, 2016, *8* (2), 154–194.

Ayres, Ian and Peter Siegelman, "Race and Gender Discrimination in Bargaining for a New Car," *American Economic Review*, 1995, pp. 304–321.

Bagues, Manuel and Berta Esteve-Volart, "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment," *Review of Economic Studies*, 2010, *77* (4), 1301–1328.

Bagues, Manuel, Mauro Sylos-Labini, and Natalia Zinovyeva, "Does the Gender Composition of Scientific Committees Matter?," *American Economic Review*, 2017, *107* (4), 1207–38.

Barrett, Jessica, Steven Gortmaker, Michael Long, Zachary Ward, Stephen Resch, Marj Moodie, Rob Carter, Gary Sacks, Boyd Swinburn, Y. Claire Wang, and Angie Cradock, "Cost Effectiveness of an Elementary School Active Physical Education Policy," *American Journal of Preventative Medicine*, 2015, *49* (1), 148–159.

Bartik, Alexander, Janet Currie, Michael Greenstone, and Christopher Knittel, "The local economic and welfare consequences of hydraulic fracturing," 2016.

Batson v. Kentucky, "476 US 79," 1986.

Becker, Gary, "Crime and punishment: An economic approach," *Journal of Political Economy*, 1968, pp. 169–217.

Bell, Brian, Anna Bindler, and Stephen Machin, "Crime Scars: Recessions and the Making of Career Criminals," *Review of Economics and Statistics*, Forthcoming.

Bell, Brian, Francesco Fasani, and Stephen Machin, "Crime and immigration: Evidence from large immigrant waves," *Review of Economics and statistics*, 2013, *21* (3), 1278–1290.

Bernasco, Wim, Thomas de Graaff, Jan Rouwendal, and Wouter Steenbeek, "Social interactions and crime revisited: An investigation using individual offender data in Dutch neighborhoods," *Review of Economics and Statistics*, 2017, *99* (4), 622–636.

Besley, Timothy and Anne Case, "Unnatural experiments? Estimating the incidence of endogenous policies," *The Economic Journal*, 2000, *110* (467), 672–694.

Bindler, Anna and Randi Hjalmarsson, "The Persistence of the Criminal Justice Gender Gap: Evidence from 200 Years of Judicial Decisions," 2017. Working paper.

Breda, Thomas and Son Thierry Ly, "Professors in Core Science Fields Are Not Always Biased against Women: Evidence from France," *American Economic Journal: Applied Economics*, 2015, *7* (4), 53–75.

Butcher, Kristin and Anne Morrison Piehl, "Why are immigrants' incarceration rates so low? Evidence on selective immigration, deterrence, and deportation," 2007. NBER Working Paper 13229.

Butcher, Kristin, Kyung Park, and Anne Morrison Piehl, "Comparing Apples to Oranges: Differences in WomenâĂŹs and MenâĂŹs Incarceration and Sentencing Outcomes," *Journal of Labor Economics*, 2017, *35* (S1), S201–S234.

Butt, Joanne, Robert Weignberg, Jeff Breckon, and Randal Claytor, "Adolescent Physical Activity Participation and Motivational Determinants Across Gender, Age, and Race," *Journal of Physical Activity and Health*, 2011, *8*, 1074–1083.

Calonico, Sebastian, Matias Cattaneo, Max Farrell, and Rocıo Titiunik, "rdrobust: Software for Regression Discontinuity Designs," Technical Report, University of Michigan 2016.

Card, David and Laura Giulliano, "Peer Effects and Multiple Equilibria in the Risky Behavior of Friends," *Review of Economics and Statistics*, 2013, *95* (4), 1130–1149.

Carlson, Susan, Janet Fulton, Sarah Lee, L. Michele Maynard, David Brown, Harold Kohl III, and William Dietz, "Physical Education and Academic Achievement in Elementary School: Data from the Early Childhood Longitudinal Study," *American Journal of Public Health*, 2008, *98* (4), 721–727.

Case, Ann and Angus Deaton, "Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century," *PNAS*, 2015, *112* (49), 15078–15083.

Case, Ann and Angus Deaton, "Mortality and morbidity in the 21st century," *Brookings Pap Econ Act*, 2017, pp. 397–476.

Cawley, John, Chad Meyerhoefer, and David Newhouse, "The Impact of State Physical Education Requirements on Youth Physical Activity and Overweight," *Health Economics*, 2007, *16* (12), 1287–1301.

Cawley, John, David Frisvold, and Chad Meyerhoefer, "The Impact of Physical Education on Obesity Among Elementary School Children," *Journal of Health Economics*, 2013, *32* (4), 743–755.

Centers for Disease Control and Prevention, "Physical Education and Physical Activity: Results from the School Health Policies and Programs Study 2006," 2007.

Centers for Disease Control and Prevention, "The Association Between School-Based Physical Activity, Including Physical Education, and Academic Performance," 2010. Accessed 13-November-2017 at `https://www.cdc.gov/healthyyouth/health_and_academics/pdf/pa-pe_paper.pdf`.

Centers for Disease Control and Prevention, "Childhood Obesity Facts," 2016. Accessed 1-July-2016 at `www.cdc.gov/healthyschools/obesity/facts.htm`.

Chalfin, Aaron, "The Long-Run Effect of Mexican Immigration on Crime in US Cities: Evidence from Variation in Mexican Fertility Rates," *American Economic Review Papers & Proceedings*, 2015, *105* (5), 220–25.

Cohen, Alma and Crystal Yang, "Judicial Politics and Sentencing Decisions," *American Economic Journal: Economic Policy*, Forthcoming.

Cook, Philip, *The Supply and Demand of Criminal Opportunities*, Vol. 7, Chicago:University of Chicago Press, 1986.

Crooks, Eds, "The US Shale Revolution," 2015. Financial Times Accessed 1-August-2018 `https://www.ft.com/content/2ded7416-e930-11e4-a71a-00144feab7de`.

Cutler, David, Edward Glaser, and Jesse Shapiro, "Why Have Americans Become More Obese?," *Journal of Economic Perspectives*, 2003, *17* (3), 93–117.

Dahl, Gordon and Enrico Moretti, "The Demand for Sons," *Review of Economic Studies*, 2008, *75* (4), 1085–1120.

Datar, Aashlesha and Roland Sturm, "Physical Education in Elementary School and Body Mass Index: Evidence from the Early Childhood Longitudinal Study," *American Journal of Public Health*, 2004, *94* (9), 1501–1506.

De Paola, Maria and Vincenzo Scoppa, "Gender Discrimination and EvaluatorsâĂŹ Gender: Evidence from Italian Academia," *Economica*, 2015, *82* (325), 162–188.

Depew, Briggs, Ozkan Eren, and Naci Mocan, "Judges, Juveniles, and In-Group Bias," *Journal of Law and Economics*, 2017, *60* (2), 209–239.

di Tella, Rafael, "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack," *American Economic Review*, 2004, *39* (1), 65–73.

Dills, Angela, Hillary Morgan, and Kurt Rotthoff, "Recess, Physical Education, and Elementary School Student Outcomes," *Economics of Education Review*, 2011, *30*, 889–900.

Dix-Carneiro, Rafeal, Rodrigo Soares, and Gabriel Ulyssea, "Economic Shocks and Crime: Evidence from the Brazilian Trade Liberalization," *American Economic Journal: Applied Economics*, Forthcoming.

Eren, Ozkan and Naci Mocan, "Emotional Judges and Unlucky Juveniles," *American Economic Journal: Applied Economics*, 2018, *10* (3), 171–205.

Evans, William and Julie Topoleski, "The social and economic impact of Native American casinos," 2002. National Bureau of Economic Research Working Paper.

Fetzer, Thiemo, "Fracking growth," 2014.

Feyrer, James, Erin Mansur, and Bruce Sacerdote, "Geographic dispersion of economic shocks: Evidence from the fracking revolution," *American Economic Review*, 2017, *107* (4), 1313–34.

Finkelstein, Eric, Ian Fiebelkorn, and Guijing Wang, "National Medical Spending Attributable to Overweight and Obesity: How Much, and Who's Paying," *Health Affairs*, 2003, *W3*, 219–226.

Finkelstein, Eric, Justin Trogdon, Joel Cohen, and William Dietz, "Annual Medical Spending Attributable to Obesity: Payer-and Service-Specific Estimates," *Health Affairs*, 2009, *28* (5), w822–w831.

Flanagan, Francis, "Race, Gender, and Juries: Evidence from North Carolina," *Journal of Law and Economics*, 2018, *61* (2), 189–214.

Freedman, Matthew and Emily Owens, "Your Friends and Neighbors: Localized Economics Development and Criminal Activity," *Review of Economics and Statistics*, 2016, *98* (2), 233–253.

Galbiati, Roberto, Aurélie Ouss, and Arnaud Philippe, "Jobs, News and Re-offending after Incarceration," 2018. Working Paper.

Gallup News Service, "Record-High Support for Legalizing Marijuana use in U.S.," 2017. Accessed 8-March-2018 at `http://news.gallup.com/poll/221018/record-high-support-legalizing-marijuana.aspx`.

George, Tracey, "Court Fixing," *Arizona Law Review*, 2001, *43* (1), 9–62.

Gibbs, Chloe, Jens Ludwig, and Douglas Miller, "Does Head Start Do Any Lasting Good?," 2011. NBER Working Paper No. 17452.

Gion, Cody, Kent McIntosh, and Robert Horner, "Patterns of Minor Office Discipline Referrals in Schools Using SWIS," 2014. Accessed 29-November-2018 at `https://www.pbis.org/Common/Cms/files/pbisresources/EvalBrief_May2014.pdf`.

Gittings, R. Kaj and Travis Roach, "Who really benefits from a resource boom? Evidence from the Marcellus and Utica Shale Plays," 2018. Working Paper.

Glaeser, Edward, Bruce Sacerdote, and Jose Scheinkman, "Crime and Social Interactions," *Quarterly Journal of Economics*, 1996, *11* (2), 507–548.

Goldin, Claudia and Cecilia Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 2000, *90* (4), 715–741.

Gould, Eric, Bruce Weinberg, and David Mustard, "Crime rates and local labor market opportunities in the United States: 1979–1997," *Review of Economics and statistics*, 2002, *84* (1), 45–61.

Grieco, Justin, "BRAC and crime: examining the effects of an installationâĂŹs closure on local crime," 2017. Working Paper.

Grinols, Earl and David Mustard, "Casinos, Crime, and Community Costs," *Review of Economics and Statistics*, 2006, *88* (1), 28–45.

Guerra, Paulo, Moacyr Cuce Nobre, Jonas Cardoso da Silverira, and José de Aguiar Carrazedo Taddei, "The Effect of School-Based Physical Activity Interventions on Body Mass Index: A Meta-Analysis of Randomized Trials," *Clinics (Sao Paolo)*, 2013, *68* (9), 1263–1273.

Guo, Jeff, Terrance Wade, Wei Pan, and Kathryn Keller, "School-Based Health Centers: Cost-Benefit Analysis and Impact on Health Care Disparities," *American Journal of Public Health*, 2010, *100* (9), 1617–1623.

Harbaugh, William, Naci Mocan, and Michael Visser, "Theft and deterrence," *Journal of Labor Research*, 2013, *34* (4), 389–407.

Hoxby, Caroline, Sonali Murarka, and Jenny Kang, "How New York City's Charter Schools Affect Achivement, August 2009 Report," 2009.

Huffington Post/YouGov, "Survey," 2013. Accessed 10-March-2018 at `http://big.assets.huffingtonpost.com/toplines_sentences_0813142013.pdf`.

Institute of Medicine, "Accelerating Progress in Obesity Prevention: Solving the Weight of the Nation. National Academies Press, Washington, DC.," 2012. Accessed 4-June-2018 at `https://www.ncbi.nlm.nih.gov/pubmed/24830053`.

Institute of Medicine, "Educating the Student Body: Taking Physical Activity and Physical Education to School," 2013. Accessed 2-October-2017 at `https://www.ncbi.nlm.nih.gov/pubmed/24851299`.

Jacob, Brian and Lars Lefgren, "Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime," *American Economic Review*, 2003, *93* (5), 1560–1577.

Janssen, Ian, Wendy Craig, William Boyce, and William Picket, "Associations Between Overweight and Obesity with Bullying Behaviors in School-Aged Children," *Pediatrics*, 2004, *113* (5), 1187–1194.

J.E.B. v. Alabama, "511 US 127," 1994.

Jensen, Chad, Christopher Cushing, and Allison Elledge, "Associations Between Teasing, Quality of Life, and Physical Activity Among Preadolescent Children," *Journal of Pediatric Psychology*, 2013, *39* (1), 65–73.

Johnson, Brian, "Judges on Trial: A Reexamination of Judicial Race and Gender Effects across Modes of Conviction," *Criminal Justice Policy Review*, 2014, *25* (2), 159–184.

Knaus, Michael, Michael Lechner, and Anne Reimers, "For Better or Worse? The Effects of Physical Education on Child Development," 2018. IZA Discussion Paper No. 11268.

Knepper, Matthew, "When the Shadow Is the Substance: Judge Gender and the Outcomes of Workplace Sex Discrimination Cases," *Journal of Labor Economics*, 2018, *36* (3), 623–664.

Komarek, Timothy, "Crime and Natural Resource Booms: Evidence from Unconventional Natural Gas Production," 2017. Working Paper.

Latner, Janet and Albert Stunkard, "Getting Worse: The Stigmatization of Obese Children," *Obesity Research*, 2003, *11* (3), 452–456.

Lavy, Victor, "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment," *Journal of Public Economics*, 2008, *92* (10-11), 2083–2105.

Lee, David and David Card, "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 2008, *142*, 655–674.

Let's Move, "The Epidemic of Childhood Obesity," 2016. Accessed 02-July-2016 at `http://www.letsmove.gov/learn-facts/epidemic-childhood-obesity`.

Ludwig, Jens and Jeffrey Kling, "Is Crime Contagious?," *Journal of Law and Economics*, 2007, *50* (3).

Machin, Stephin and Oliver Marie, "Crime and Police Resources: the Street Crime Initiative," *Journal of the European Economic Association*, 2011, *9* (4), 678–701.

Maniloff, Peter and Ralph Mastromonaco, "The local economic impacts of hydraulic fracturing and determinants of Dutch disease," *Colorado School of Mines, Division of Economics and Business Working Paper*, 2014, *8.*

McDonough Power Equipment, Inc. v. Greenwood, "464 U.S. 548," 1984.

Mejia, Daniel and Pascual Restrepo, "Crime and Conspicuous Consumption," *Journal of Public Economics*, 2016, *135*, 1–14.

Miles, Thomas and Adam Cox, "Does immigration enforcement reduce crime? Evidence from secure communities," *The Journal of Law and Economics*, 2014, *57* (4), 937–973.

Mitchell, Tara, Ryann Haw, Jeffrey Pfeifer, and Christian Meissner, "Racial Bias in Mock Juror Decision-Making: A Meta-Analytic Review of Defendant Treatment," *Law and Human Behavior*, 2005, *29* (6), 621–637.

Montolio, Daniel, "The Unintended Consequences on Crime of âĂŸPennies from HeavenâĂŹ," 2018. IDB Working Paper No. IDB-WP-666.

Morning Consult, "National Tracking Poll 160304," 2016. Accessed 10-March-2018 at `https://cdn0.vox-cdn.com/uploads/chorus_asset/file/6189021/Morning_Consult_Vox_drug_poll.0.pdf`.

Moss-Racusin, Corinne, John Dovidio, Victoria Brescoll, Mark Graham, and Jo Handelsman, "Science FacultyâĂŹs Subtle Gender Biases Favor Male Students," *Proceedings of the National Academy of Sciences*, 2012, *109* (41), 16474–16479.

Mueller-Smith, Michael, "The Criminal and Labor Market Impacts of Incarceration," *American Economic Review*, Forthcoming.

Nader, Philip, Robert Bradley, Renate Houts, and Marcy O'Brien, "Moderate-to-Vigorous Physical Activity From Ages 9 to 15 Years," *The Journal of the American Medical Association*, 2008, *300* (3), 247–352.

National Center for Education Statistics, "Indicator 7: Discipline Problems Reported by Public Schools," 2018. Accessed 29-November-2018 at `https://nces.ed.gov/programs/crimeindicators/ind_07.asp`.

National Conference of State Legislatures (NCSL), "Drug Sentencing Trends," 2016. Accessed 10-March-2018 at `http://www.ncsl.org/research/civil-and-criminal-justice/drug-sentencing-trends.aspx`.

National Organization for the Reform of Marijuana Laws (NORML), "States That Have Decriminalized," 2018. Accessed 10-March-2018 at `http://norml.org/aboutmarijuana/item/states-that-have-decriminalized`.

Neumark, David, Roy Bank, and Kyle Van Nort, "Sex Discrimination in Restaurant Hiring: An Audit Study," *Quarterly Journal of Economics*, 1996, *111* (3), 915–941.

North Dakota Labor Market Information, "Oil and Gas Producing Counties," 2018. Accessed 02-July-2018 at `https://www.ndworkforceintelligence.com/gsipub/index.asp?docid=545`.

Pate, Russell and Jennifer O'Neill, "Summary of the American Heart Association Scientific Statement: Promoting Physical Activity in Children and Youth: A Leadership Role for Schools," *The Journal of Cardiovascular Nursing*, 2008, *23*, 44–49.

Pew Research Center, "AmericaâĂŹs New Drug Policy Landscape," 2014. Accessed 10-March-2018 at `http://www.people-press.org/2014/04/02/section-1-perceptions-of-drug-abuse-views-of-drug-policies/`.

Philippe, Arnaud and Aurélie Ouss, ""No Hatred or Malice, Fear or Affection": Media and Sentencing," *Journal of Political Economy*, 2017.

Price, Joseph and Justin Wolfers, "Racial Discrimination among NBA Referees," *Quarterly Journal of Economics*, 2010, *125* (4), 1859–1887.

Raphael, Steven and Rudolf Winter-Ebmer, "Identifying the effect of unemployment on crime," *The Journal of Law and Economics*, 2001, *44* (1), 259–283.

Schanzenbach, Diane, "What Have Researchers Learned from Project STAR?," *Brookings Papers on Education Policy*, 2007, *9*, 205–228.

Schanzenbach, Max, "Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics," *Journal of Legal Studies*, 2005, *34* (1), 57–92.

Schnepel, Kevin, "Good jobs and recidivism," *The Economic Journal*, 2017, *128* (608), 447–469.

Shayo, Moses and Asaf Zussman, "Judicial Ingroup Bias in the Shadow of Terrorism," *Quarterly Journal of Economics*, 2011, *126* (3), 1447–1484.

Spenkuch, Jörg, "Understanding the impact of immigration on crime," *American law and economics review*, 2013, *16* (1), 177–219.

Stanley, Tom and Stephen Jarrell, "Gender Wage Discrimination Bias? A Meta-Regression Analysis," *Journal of Human Resources*, 1998, pp. 947–973.

Steffensmeier, Darrell and Chris Hebert, "Women and Men Policymakers: Does the Judge's Gender Affect the Sentencing of Criminal Defendants?," *Social Forces*, 1999, *77* (3), 1163–1196.

Texas Education Agency, "Evaluation of the Texas Fitness Now Grant Program: 2007-08 to 2009-10 School Years," 2011. Accessed 01-July-2016 at `http://tea.texas.gov/WorkArea/DownloadAsset.aspx?id=2147496810`.

Texas Education Agency, National Center on Performance Initiatives, "Texas Educator Excellence Grant (TEEG) Program: Year Three Evaluation Report," 2009. Accessed 01-June-2018 at `https://tea.texas.gov/WorkArea/DownloadAsset.aspx?id=2147490696`.

Tomporowsk, Phillip, Catherine Davis, Patricia Miller, and Jack Naglieri, "Exercise and Children's Intelligence, Cognition, and Academic Achievement," *Educational Psychology Review*, 2008, *20* (2), 111–131.

Tremarche, Pamela, Ellyn Robinson, and Louise Graham, "Physical Education and its Effect on Elementary Testing Results," *The Physical Educator*, 2007, *64* (2), 58–64.

Trudeau, François and Roy Shephard, "Physical Education, School Physical Activity, School Sports and Academic Performance," *International Journal of Behavioral Nutrition and Physical Activity*, 2008, *5* (10), 1–12.

U.S. Bureau of Economic Analysis, "Real Total Gross Domestic Product for North Dakota," 2018. Accessed 18-September-2018 `https://fred.stlouisfed.org/series/NDRGSP`.

U.S. Bureau of the Census, "Median Household Income in North Dakota," 2018. Financial Times Accessed 18-September-2018 `https://fred.stlouisfed.org/series/MEHOINUSNDA646N`.

U.S. Department of Education, "Carol M. White Physical Education Program," 2013. Accessed 23-November-2018 at https://www2.ed.gov/programs/whitephysed/funding.html.

Vollaard, Ben and Joseph Hamed, "Why the Police have and Effect on Violent Crime After All: Evidence from the British Crime Survey," *Journal of Law and Economics*, 2012, *55* (4).

von Hippel, Paul and Kyle Bradbury, "The Effects of School Physical Education Grants on Obesity, Fitness, and Academic Achievement," *Preventative Medicine*, 2015, *78*, 44–51.

Wang, Li, Quanhe Yang, Richard Lowry, and Howell Wechsler, "Economic Analysis of a School-Based Obesity Prevention Program," *Obesity*, 2003, *11* (11), 1313–1324.

Wang, Zhongmin and Alan Krupnick, "A retrospective review of shale gas development in the United States: What led to the boom?," 2013.

Waters, Elizabeth, Andrea de Silva-Sanigorski, Belinda Bedford, Tamara Brown, Karen Campbell, Yang Gao, Rebecca Armstrong, Lauren Prosser, and Carolyn Summerbell, "Interventions for Preventing Obesity in Children," *Cochrane Database of Systematic Reviews*, 2011. Accessed 2-October-2017 at http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001871.pub3/abstract.

Weber, Jeremy, "A decade of natural gas development: The makings of a resource curse?," *Resource and Energy Economics*, 2014, *37*, 168–183.

West, Jeremy, "Racial Bias in Police Investigations," 2017. Working paper.

Wilson, Riley, "Moving to Economic Opportunity: The Migration Response to the Fracking Boom," 2017. Working Paper.

Yang, Crystal, "Local labor markets and criminal recidivism," *Journal of Public Economics*, 2017, *147*, 16–29.

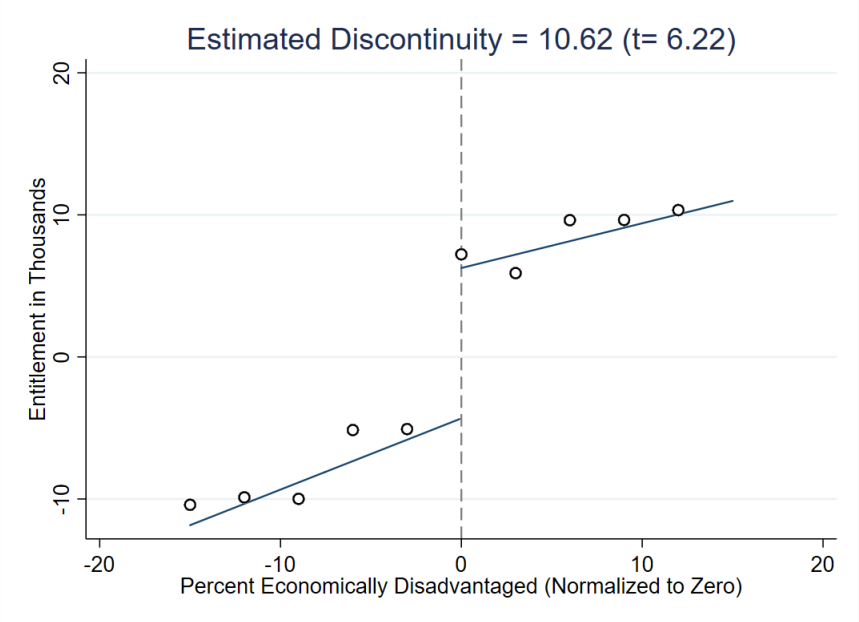
YouGov/Huffington Post, "Poll Results: Commuting Sentences," 2015. Accessed 10-March-2018 at https://today.yougov.com/news/2015/04/06/poll-results-commuting-sentences/.

APPENDIX A

FIGURES AND TABLES FOR SECTION 2

A.1 Figures

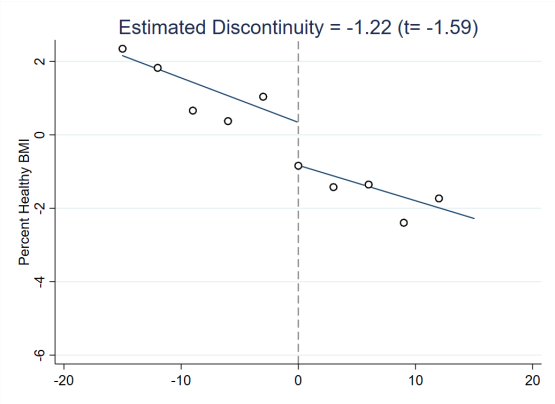
Figure A.1: The Effect of Eligibility on Funding



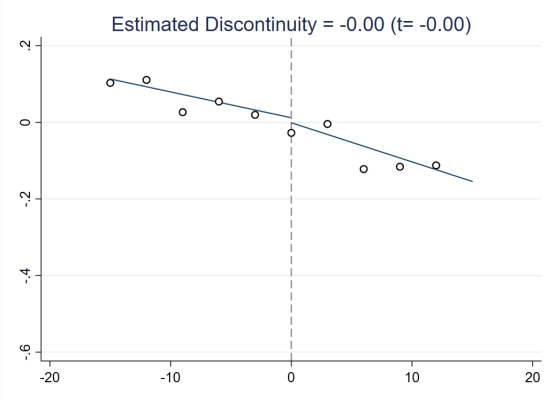
Notes: Funding data for the Texas Fitness Now (TFN) program from 2007-2011 is from the Texas Education Agency, grants division. Entitlement is calculated as the total grant allowance per school year. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. “Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes all Texas schools with students in grades 6, 7, and/or 8.

Figure A.2: The Effect of Texas Fitness Now on Physical Fitness

Percent of Students with a Healthy BMI

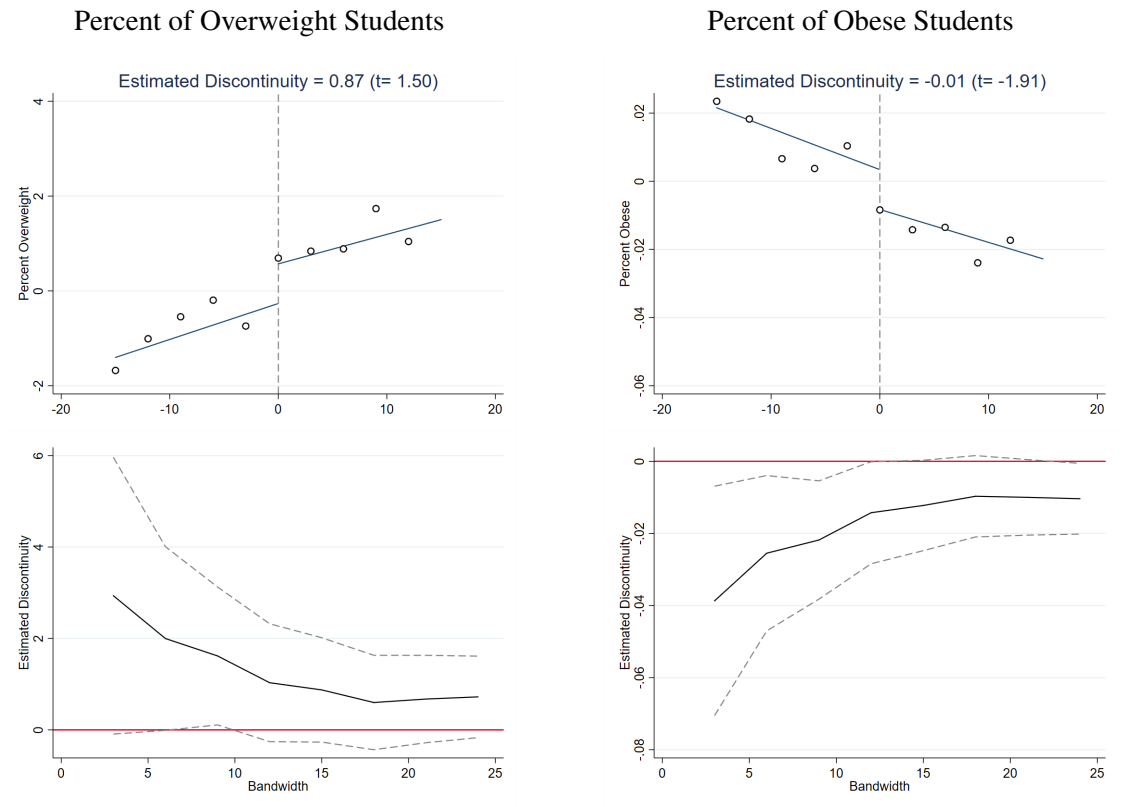


Average Number of Fitness Tests Passed



Notes: School-level data on fitness outcomes is from FITNESSGRAM© data provided by the Texas Education Agency (TEA). Each figure plots means of residuals (after differencing out year fixed effects) in 3 percentage point bins and linear fits of the outcome listed. “Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes students in Texas in grades 6, 7, and/or 8 from school years spanning 2007-2011.

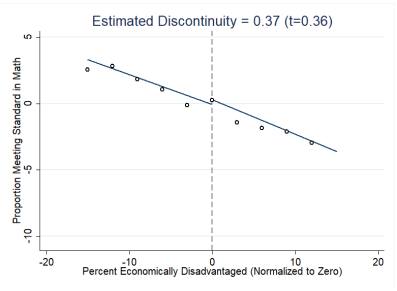
Figure A.3: Analyzing Changes in BMI for Overweight and Obese Students



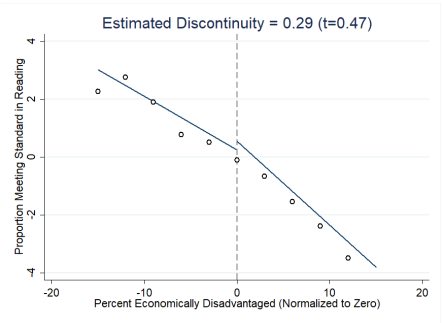
Notes: School-level data on fitness outcomes is from FITNESSGRAM© data provided by the Texas Education Agency (TEA). Each figure plots means of residuals (after differencing out year fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes students in Texas in grades 6, 7, and/or 8 from school years spanning 2007-2011.

Figure A.4: The Effect of Texas Fitness Now on Test Scores

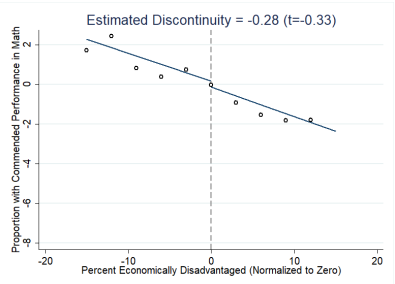
Pass Rate Math TAKS Test



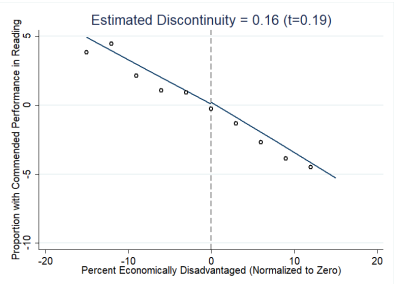
Pass Rate Reading TAKS Test



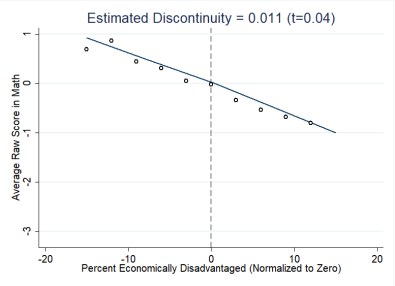
Commended Performance Math TAKS Test



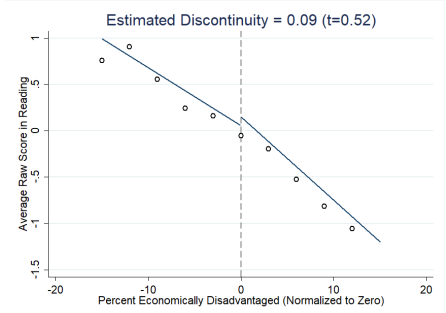
Commended Performance Reading TAKS Test



Raw Score Math TAKS Test



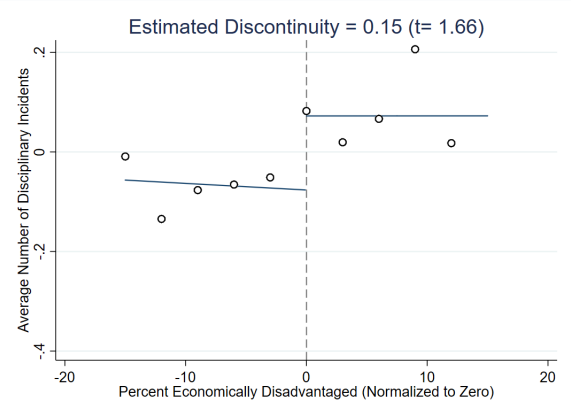
Raw Score Reading TAKS Test



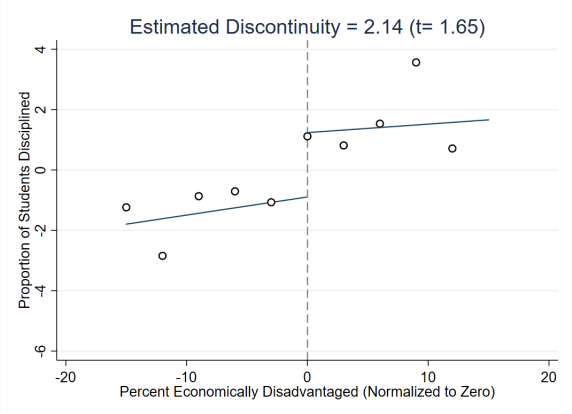
Notes: Student-level data on test scores is from the Education Research Center at UT-Austin. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. “Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and 8 from school years spanning 2007-2011.

Figure A.5: The Effect of Texas Fitness Now on Disciplinary Action

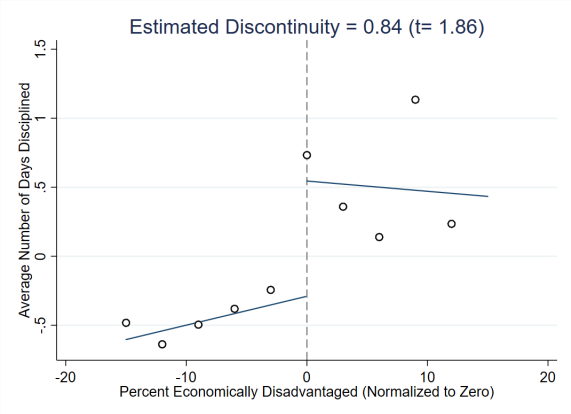
Total Disciplinary Actions



Proportion of Students with Disciplinary Action

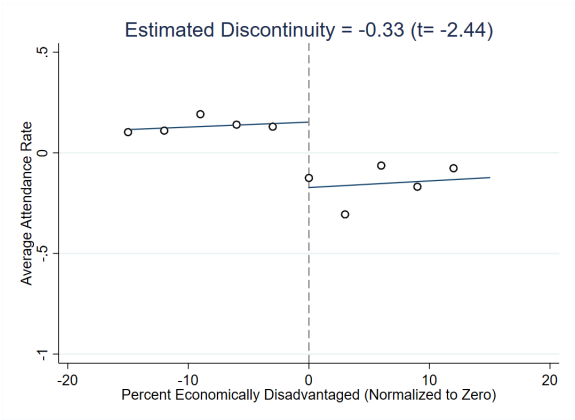


Days Suspended



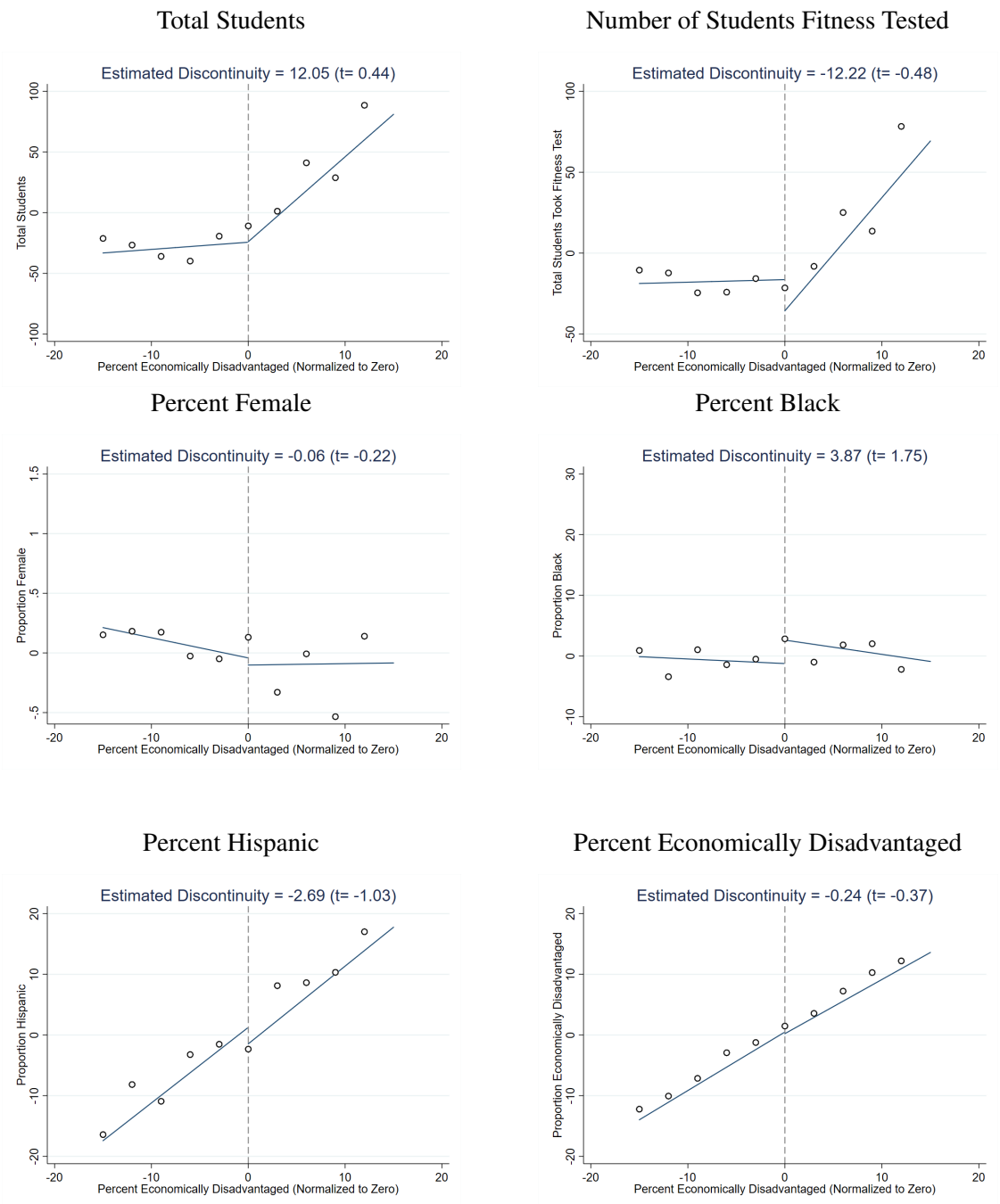
Notes: Student-level data on disciplinary outcomes is from the Education Research Center at UT-Austin. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. “Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and 8 from school years spanning 2007-2011.

Figure A.6: The Effect of Texas Fitness Now on Attendance



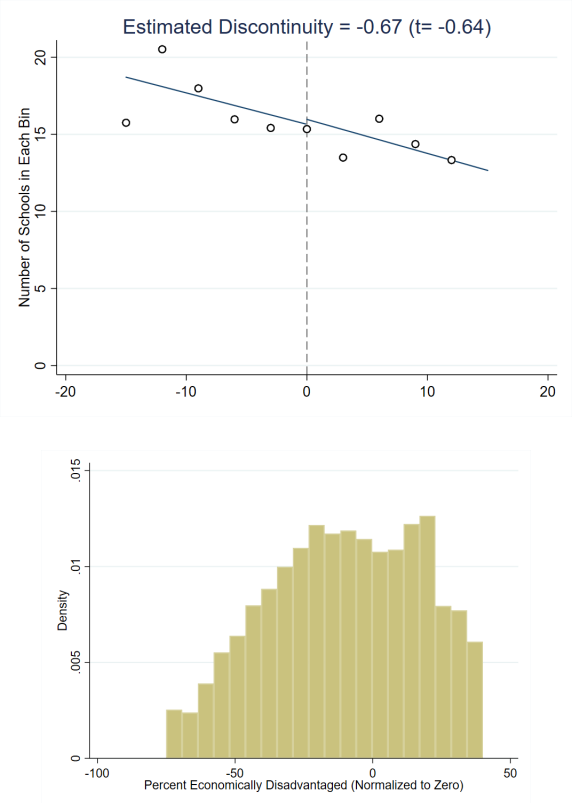
Notes: Student-level data on attendance is from the Education Research Center at UT-Austin. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and 8 from school years spanning 2007-2011.

Figure A.7: Testing Discontinuity of School Composition



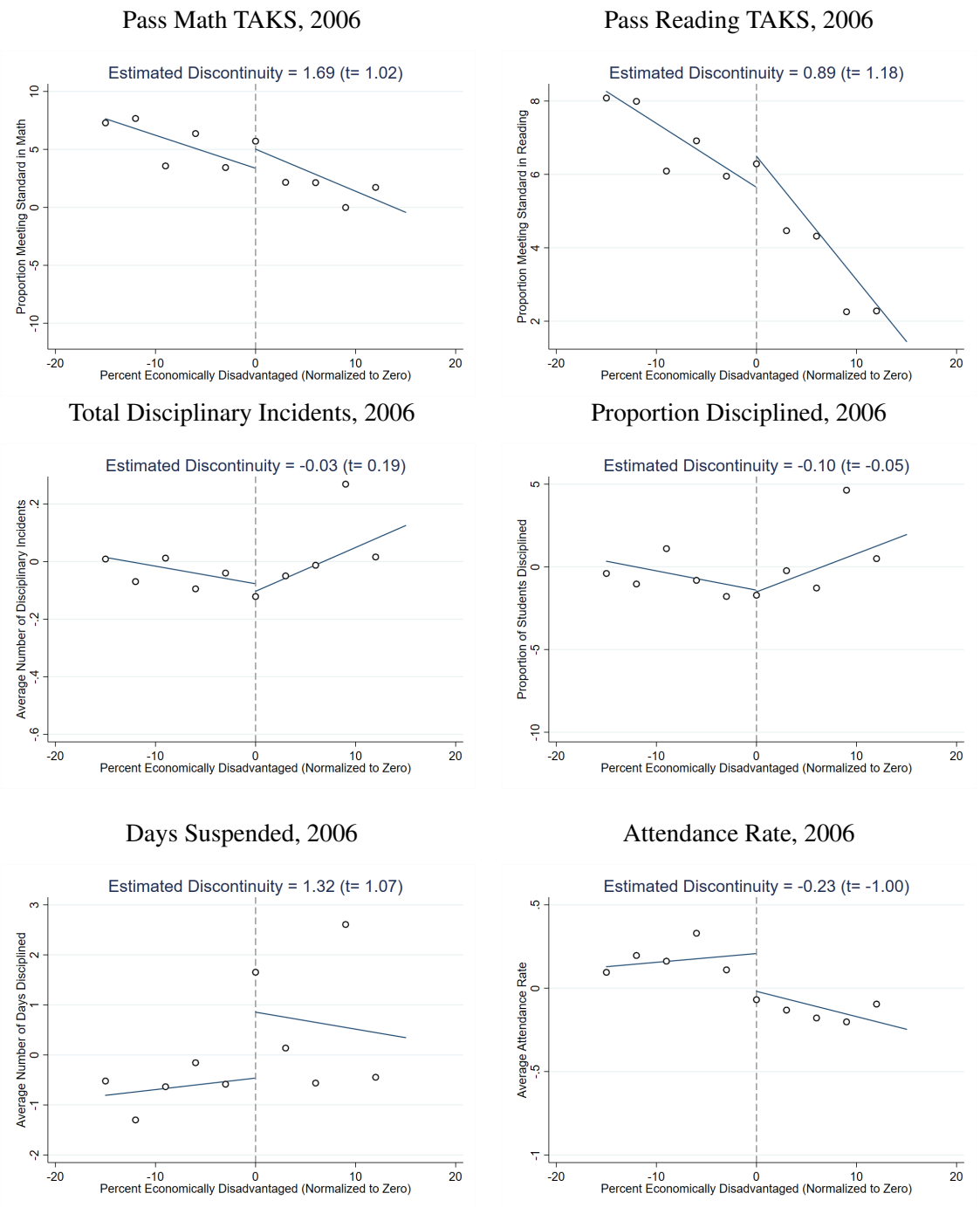
Notes: Data on school characteristics is from the Education Research Center at UT-Austin. Data on the total number of students fitness tested is from the TEA's Academic Excellence Indicator System. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and 8 from school years spanning 2007-2011.

Figure A.8: Testing the Density of Number of Bins



Notes: Data on student characteristics is from the Education Research Center at UT-Austin. Data on school characteristics is from the Texas Education Agency's Academic Excellence Indicator System. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes Texas students in grades 6, 7, and/or 8 from school years spanning 2007-2011.

Figure A.9: Testing Discontinuities in the Pre-Period



Notes: Student-level data on disciplinary action, attendance rates, and TAKS scores is from the Education Research Center at UT-Austin. Each figure plots means of residuals (after differencing out year and grade fixed effects) in 3 percentage point bins and linear fits of the outcomes listed. “Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes all students in Texas schools in grades 6, 7, and 8 from the 2006-2007 school year.

A.2 Tables

Table A.1: Texas Fitness Now Funding Schedule

School Year	ED Cutoff	Schools Eligible	Amount Granted
2007-2008	75%	605	$10,000,000
2008-2009	75%	575	$9,378,914
2009-2010	60%	981	$8,875,670
2010-2011	60%	1125	$8,500,000

Notes: Data on TFN funding and grantee awards is from the Texas Education Agency, Grants Division. ED cutoff represents the percent of economically disadvantaged students required in the previous year to be eligible for TFN funding. Total funding is approximately $37 million and average funding per school is $11,000.

Table A.2: Effects of Texas Fitness Now on Physical Fitness

	Healthy BMI			Number of Tests Passed		
	(1)	(2)	(3)	(4)	(5)	(6)
%ED > Cutoff	-2.19*	-1.42	-1.22	0.00	0.01	-0.00
	(1.19)	(0.88)	(0.77)	(0.07)	(0.05)	(0.04)
Bandwidth	6.9	12	15	5.8	12	15
Observations	1591	2769	3473	1378	2840	3555

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. School-by-grade data from the FITNESSGRAM© test for school years spanning 2007-2011 is from the Texas Education Agency. Each coefficient is generated by a separate regression of Equation 1 using the listed fitness outcome as the dependent variable, controlling for year fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction. The sample includes Texas students in grades 6, 7, or 8.

Table A.3: Effects of Texas Fitness Now on Standardized Test Scores

	Math TAKS			Reading TAKS		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Pass Test						
%ED > Cutoff	0.005	0.006	0.004	0.011	0.004	0.003
	(0.008)	(0.011)	(0.010)	(0.012)	(0.007)	(0.006)
Bandwidth	8.9	12	15	11.2	12	15
Observations	737,503	1,002,403	1,289,442	923,137	1,002,373	1,289,364
Panel B. Commended Performance						
%ED > Cutoff	0.008	0.001	-0.003	0.0003	0.002	0.002
	(0.009)	(0.010)	(0.008)	(0.010)	(0.010)	(0.008)
Bandwidth	10.9	12	15	8.2	12	15
Observations	893,230	1,002,403	1,289,442	674,118	1,002,373	1,289,364
Panel C. Raw Score						
%ED > Cutoff	0.152	0.090	0.011	0.216	0.172	0.093
	(0.248)	(0.315)	(0.278)	(0.326)	(0.206)	(0.178)
Bandwidth	8.0	12	15	11.0	12	15
Observations	663,142	999,023	1,285,172	905,681	998,993	1,285,094

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Student-level testing data for school years spanning 2007-2011 is from Education Research Center at UT-Austin. Each coefficient is generated by a separate regression of Equation 1 using the listed academic performance outcome as the dependent variable, controlling for year and grade fixed effects. A student passes an exam if they meet the standards for the test for that year. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction. The sample includes all students in Texas students in grades 6, 7, or 8.

Table A.4: Effects of Texas Fitness Now on Disciplinary Action

	Total Disciplinary Incidents			**Proportion of Students Disciplined**			**Total Days Suspended**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
%ED > Cutoff	0.080	0.075	0.149*	0.005	0.012	0.021*	0.703	0.616	0.836*
	(0.133)	(0.101)	(0.090)	(0.019)	(0.015)	(0.013)	(0.580)	(0.541)	(0.459)
Bandwidth	7.8	12	15	7.4	12	15	10.1	12	15
Observations	656,604	1,010,648	1,299,744	624,046	1,010,648	1,299,744	832,261	1,010,648	1,299,744

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Student-level data for school years spanning 2007-2011 is from Education Research Center at UT-Austin. Each coefficient is generated by a separate regression of Equation 1 using the listed discipline outcome as the dependent variable, controlling for year and grade fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in a given year. The sample includes Texas students in grades 6, 7, or 8.

Table A.5: Effects of Texas Fitness Now on Attendance

	(1)	(2)	(3)
%ED $>$ Cutoff	-0.002	-0.003**	-0.003**
	(0.002)	(0.002)	(0.001)
Bandwidth	8.9	12	15
Observations	750,912	1,008,485	1,297,023

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Student-level data for school years spanning 2007-2011 is from Education Research Center at UT-Austin. Each coefficient is generated by a separate regression of Equation 1, controlling for year and grade fixed effects. Student-level attendance rates are calculated by dividing the total number of days students were present by the total number of school days. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in a given year. The sample includes Texas students in grades 6, 7, or 8.

Table A.6: Testing Alternative Specifications

	Linear Fit (1)	Quad Fit (2)	Cubic Fit (3)	Triangular Kernel (4)
Panel A. Pass Math TAKS				
%ED > Cutoff	0.003	0.005	0.009	0.004
	(0.006)	(0.010)	(0.012)	(0.007)
Observations	1,289,364	1,289,364	1,289,364	1,289,364
Panel B. Pass Reading TAKS				
%ED > Cutoff	0.004	0.013	0.027	0.008
	(0.005)	(0.009)	(0.010)	(0.001)
Observations	1,289,442	1,289,442	1,289,442	1,289,442
Panel C. Total Disciplinary Incidents				
%ED > Cutoff	0.149*	0.012	0.099	0.092
	(0.069)	(0.127)	(0.148)	(0.010)
Observations	1,299,744	1,299,744	1,299,744	791,258
Panel D. Proportion Disciplined				
%ED > Cutoff	0.0214*	0.002	0.012	0.013
	(0.013)	(0.021)	(0.028)	(0.015)
Observations	1,299,744	1,299,744	1,299,744	1,299,744
Panel E. Days Suspended				
%ED > Cutoff	0.836*	0.412	0.752	0.666
	(0.451)	(0.670)	(0.746)	(0.512)
Observations	1,299,744	1,299,744	1,299,744	1,299,744
Panel F. Attendance Rate				
%ED > Cutoff	-0.003**	-0.002	0.001	-0.003*
	(0.001)	(0.002)	(0.002)	(0.001)
Observations	1,297,023	1,297,023	1,297,023	1,297,023

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Individual-level data on Texas middle school students from 2007-2011 is from the Education Research Center at UT-Austin. Each coefficient is generated by a separate regression of Equation 1 using the listed outcome as the dependent variable. Each regression includes year and grade fixed effects and reports results from a full one-sided bandwidth of 15. Column 1 replicates the baseline results for comparison. Columns 2 and 3 allow for the days from the cutoff to vary quadratically and cubically (in addition to on either side of the threshold), respectively. Column 4 fits the model using a triangular kernel instead of uniform kernel. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction.

A.3 Additional Results

Figure A.10: Healthy Fitness Zone Standards

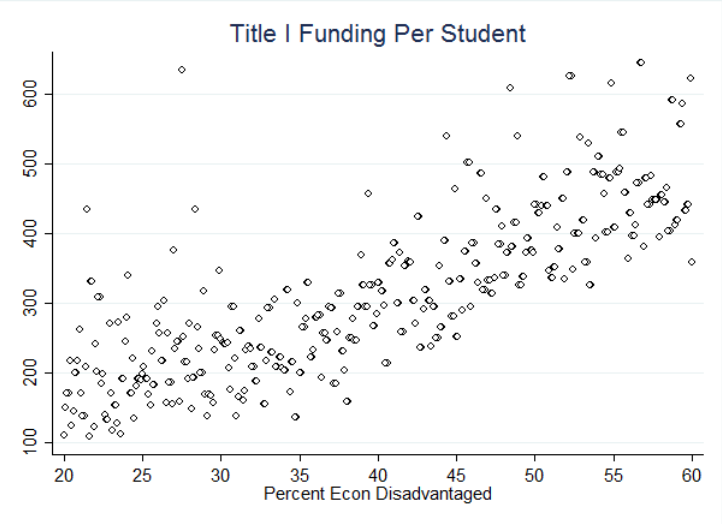
GIRLS

	Aerobic Capacity VO_{2max} (ml/kg/min) PACER, One Mile Run & Walk Test			Percent Body Fat				Body Mass Index			
	NI-Health Risk	NI	HFZ	Very Lean	HFZ	NI	NI-Health Risk	Very Lean	HFZ	NI	NI-Health Risk
5				≤9.7	9.8-20.8	20.9	≥28.4	≤13.5	13.6-16.7	16.8	≥17.3
6	*Completion of test. Lap count*			≤9.8	9.9-20.8	20.9	≥28.4	≤13.4	13.5-17.0	17.1	≥17.7
7	*or time standards not*			≤10.0	10.1-20.8	20.9	≥28.4	≤13.4	13.5-17.5	17.6	≥18.3
8	*recommended.*			≤10.4	10.5-20.8	20.9	≥28.4	≤13.5	13.6-18.2	18.3	≥19.1
9				≤10.9	11.0-22.6	22.7	≥30.8	≤13.7	13.8-18.9	19.0	≥20.0
10	≤37.3	37.4-40.1	≥40.2	≤11.5	11.6-24.3	24.4	≥33.0	≤14.0	14.1-19.5	19.6	≥21.0
11	≤37.3	37.4-40.1	≥40.2	≤12.1	12.2-25.7	25.8	≥34.5	≤14.4	14.5-20.4	20.5	≥21.9
12	≤37.0	37.1-40.0	≥40.1	≤12.6	12.7-26.7	26.8	≥35.5	≤14.8	14.9-21.2	21.3	≥22.9
13	≤36.6	36.7-39.6	≥39.7	≤13.3	13.4-27.7	27.8	≥36.3	≤15.3	15.4-22.0	22.1	≥23.8
14	≤36.3	36.4-39.3	≥39.4	≤13.9	14.0-28.5	28.6	≥36.8	≤15.8	15.9-22.8	22.9	≥24.6
15	≤36.0	36.1-39.0	≥39.1	≤14.5	14.6-29.1	29.2	≥37.1	≤16.3	16.4-23.5	23.6	≥25.4
16	≤35.8	35.9-38.8	≥38.9	≤15.2	15.3-29.7	29.8	≥37.4	≤16.8	16.9-24.1	24.2	≥26.1
17	≤35.7	35.8-38.7	≥38.8	≤15.8	15.9-30.4	30.5	≥37.9	≤17.2	17.3-24.6	24.7	≥26.7
>17	≤35.3	35.4-38.5	≥38.6	≤16.4	16.5-31.3	31.4	≥38.6	≤17.5	17.6-25.1	25.2	≥27.2

	Curl-up # *completed*	Trunk Lift *inches*		90^0 Push-up # completed	Modified Pull-up # completed	Flexed Arm Arm Hang seconds	Back Saver Sit & Reach** *inches*	Shoulder Stretch
5	≥2	6	12	≥3	≥2	≥2	9	Healthy Fitness Zone = Touching fingertips together behind the back on both right and left sides
6	≥2	6	12	≥3	≥2	≥2	9	
7	≥4	6	12	≥4	≥3	≥3	9	
8	≥6	6	12	≥5	≥4	≥3	9	
9	≥9	6	12	≥6	≥4	≥4	9	
10	≥12	9	12	≥7	≥4	≥4	9	
11	≥15	9	12	≥7	≥4	≥6	10	
12	≥18	9	12	≥7	≥4	≥7	10	
13	≥18	9	12	≥7	≥4	≥8	10	
14	≥18	9	12	≥7	≥4	≥8	10	
15	≥18	9	12	≥7	≥4	≥8	12	
16	≥18	9	12	≥7	≥4	≥8	12	
17	≥18	9	12	≥7	≥4	≥8	12	
17+	≥18	9	12	≥7	≥4	≥8	12	

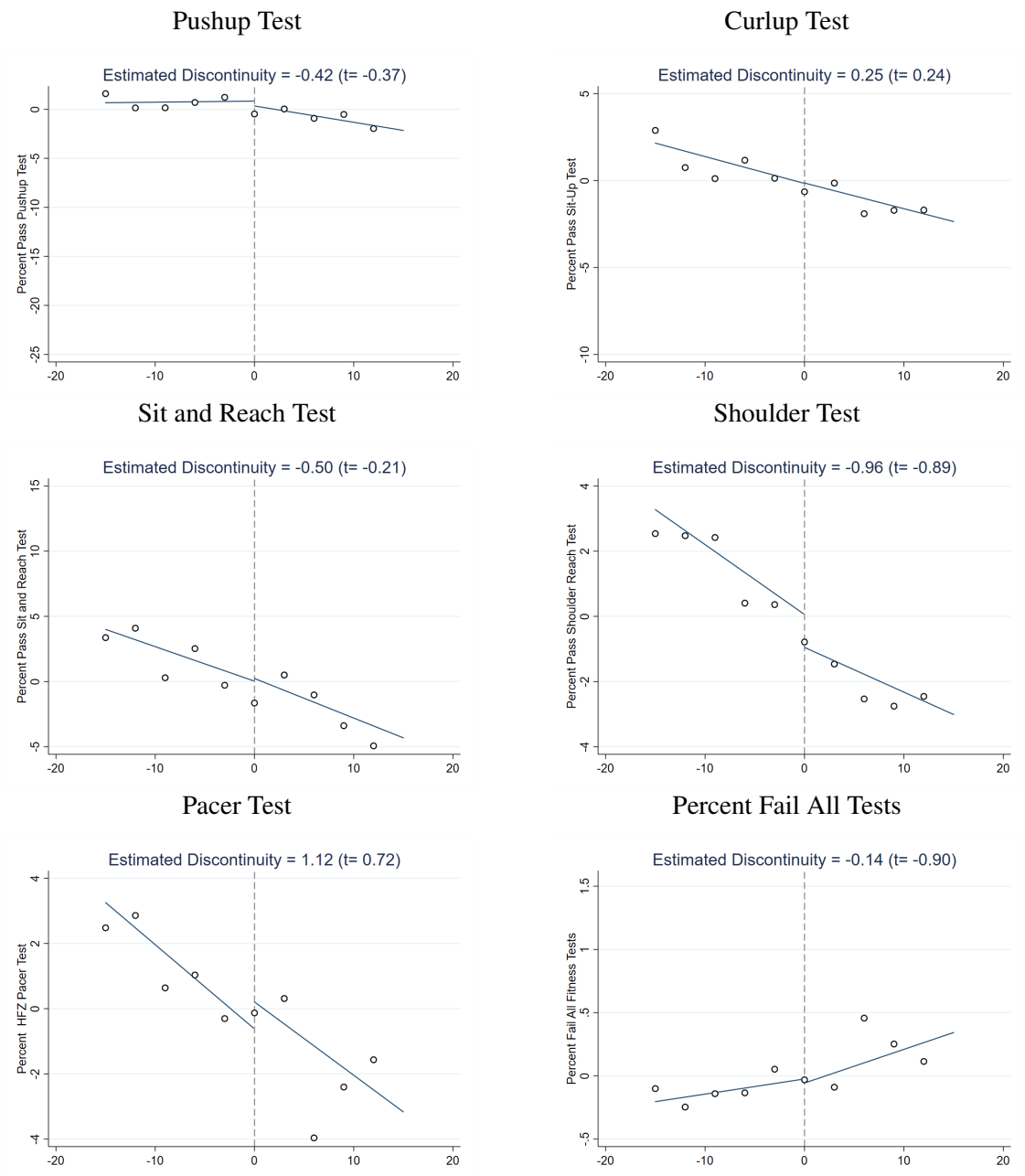
Notes: Data on FITNESSGRAM© standards for Healthy Fitness Zone are from the Cooper Institute. See `http://www.cooperinstitute.org/healthyfitnesszone` for more information.

Figure A.11: Title 1 Funding by Percent Economically Disadvantaged



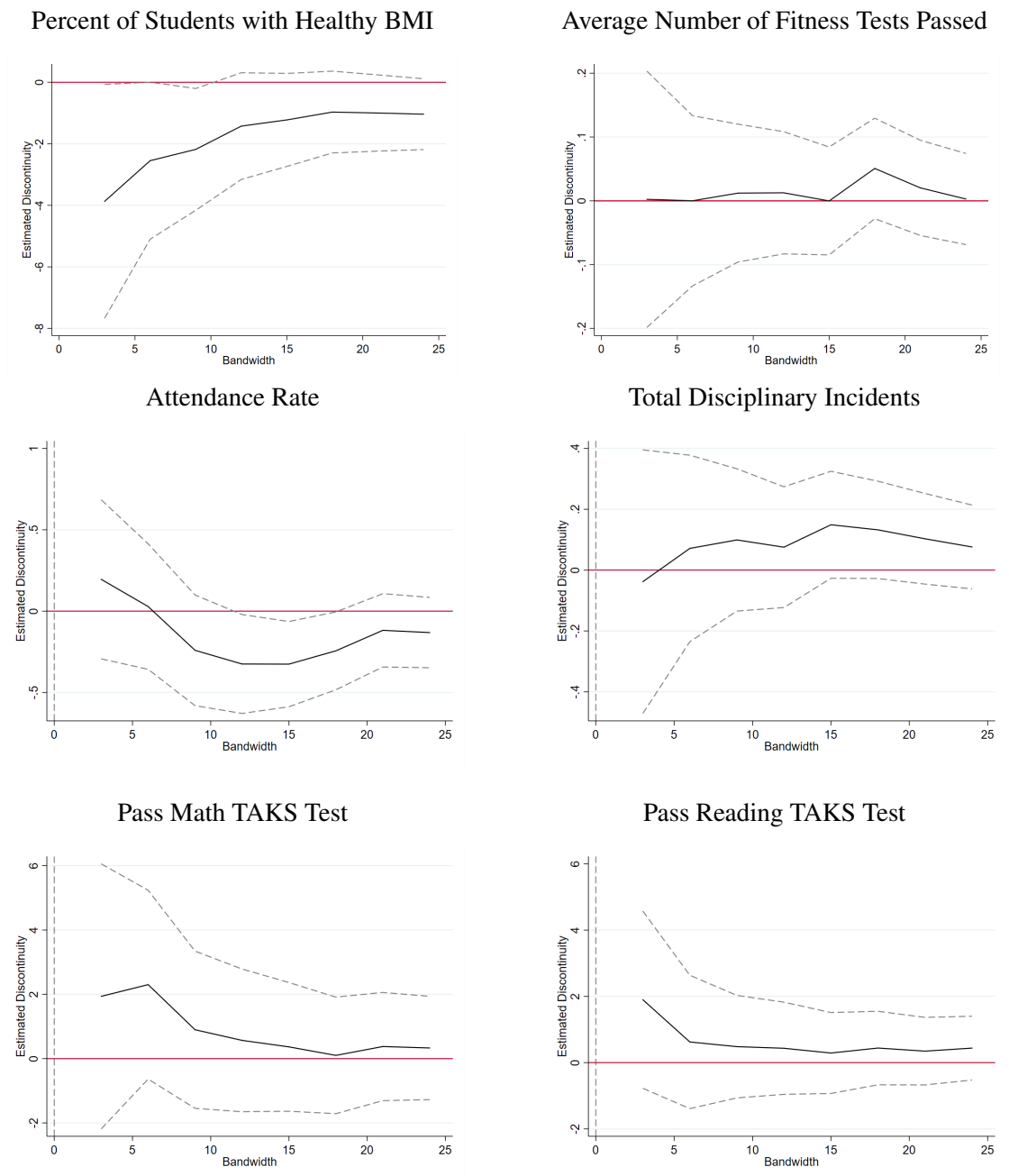
Notes: Data on district-level Title 1 funding by funding source are from the Texas Education Agency Public Education Information Management Systems Reports. Title I funds are aimed at schools with at least 40% economically disadvantaged students.

Figure A.12: The Effect of Texas Fitness Now on Physical Fitness



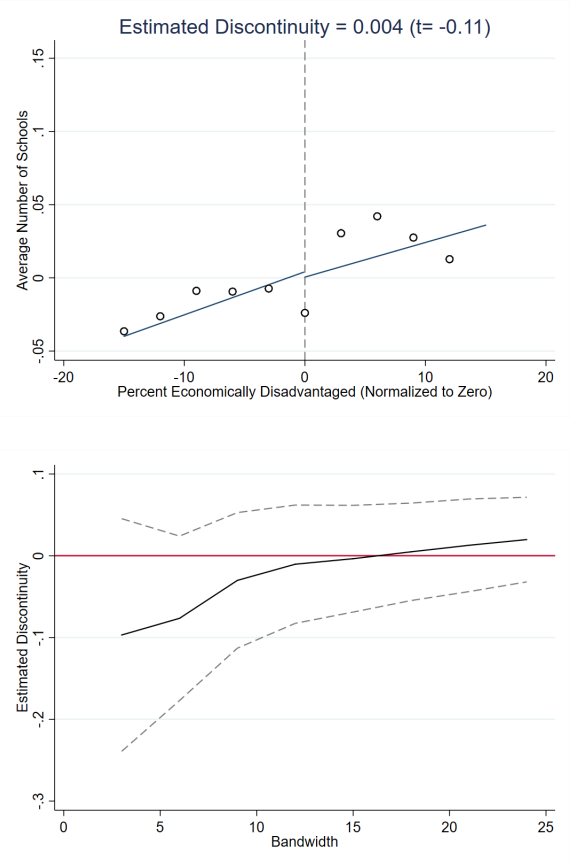
Notes: School-level data on fitness outcomes is from FITNESSGRAM© data provided by the Texas Education Agency (TEA). Each figure plots means of residuals (after differencing out year fixed effects) in 3 percentage point bins and linear fits of the outcome listed. "Estimated Discontinuity" reports estimates from a linear regression, specified in Equation 1, using uniform kernel weights and allowing the slopes to vary on each side of the threshold. The sample includes students in Texas in grades 6, 7, and/or 8 from school years spanning 2007-2011.

Figure A.13: Effect of Varying Bandwidth on Estimates



Notes: School-level data on BMI and physical fitness is from FITNESSGRAM© data provided by the Texas Education Agency (TEA). Individual-level data on test scores, discipline, and attendance is from the Education Research Center at UT-Austin. Each panel reports estimates and their corresponding 95% confidence intervals from linear regressions, using uniform kernel weights and allowing the slopes to vary on each side of the threshold, for a range of different bandwidths. The sample includes Texas students in grades 6, 7, or 8 from school years spanning 2007-2011.

Figure A.14: Testing Student Attrition



Notes: Individual-level data on school enrollment is from the Education Research Center at UT-Austin. The sample includes Texas students in grades 6, 7, and 8 from school years 2007-2008 to 2010-2011.

Table A.7: Summary Statistics

	Mean	St. Dev.	Min	Max
School Characteristics				
Total Number of Students Enrolled	515	337	1	1,816
Amount Entitled by Texas Fitness Now Grant	5,024.02	9,706.64	0	62,442
Percent Economically Disadvantaged	70.6	45.6	0	100
Percent Female	48.6	50.0	0	100
Percent White	22.4	26.18	0	100
Percent Black	17.0	37.6	0	100
Percent Hispanic	57.4	49.5	0	100
Charter School	0.04	0.19	0	1
Health and Fitness Outcomes				
Percent Healthy BMI	63.41	12.06	0	100
Percent Healthy Body Fat	73.77	21.50	0	100
Percent Pass Pacer Test	58.01	22.62	0	100
Percent Complete Mile Run	60.82	23.61	0	100
Percent Pass Push-Up Test	73.57	16.67	0	100
Percent Pass Curl up Test	79.49	16.68	0	100
Percent Pass Sit and Reach Test	64.56	25.00	0	100
Percent Pass Shoulder Test	72.44	13.66	0	100
Percent Pass All Fitness Tests	22.87	15.77	0	88
Percent Fail All Fitness Tests	1.08	2.52	0	71
Academic Outcomes				
Math TAKS Passing Rate	0.71	0.45	0	1
Reading TAKS Passing Rate	0.83	0.38	0	1
Math TAKS Commended Rate	0.20	0.40	0	1
Reading TAKS Commended Rate	0.33	0.47	0	1
Math TAKS Raw Score	31.70	11.91	0	50
Reading TAKS Raw Score	35.24	11.74	0	48
Total Disciplinary Incidents	0.96	0.44	0	97
Proportion of Students Disciplined	0.27	2.65	0	1
Total Days Suspended	3.54	15.59	0	910
Attendance Rate	0.96	0.05	0.01	1

Notes: Individual-level data on student characteristics and academic outcomes, including economically disadvantaged status, race, ethnicity, test scores, discipline, and attendance are from the Education Research Center at UT-Austin. Data on fitness outcomes are from the standardized fitness testing program, FITNESSGRAM©, are from the Texas Education Agency (TEA). Texas Fitness Now grant entitlements data are from the publicly-available list of grantee awards provided by the TEA. Entitlements per student for each school are calculated using the total amount of funding divided by enrollment. The sample includes Texas students in grades 6, 7, or 8 from school years spanning 2007-2011.

Table A.8: Effects of Funding Cuts on Physical Fitness- Females

	Healthy BMI			Number of Tests Passed		
	(1)	(2)	(3)	(4)	(5)	(6)
%ED > Cutoff	-0.92	-0.59	-0.31	-0.28	-0.12	0.05
	(1.56)	(1.14)	(1.03)	(0.20)	(0.17)	(0.15)
Bandwidth	7.6	12	15	7.7	12	15
Observations	1455	2762	3454	1774	2762	3454
	Pacer Test			Mile Run		
%ED > Cutoff	1.49	1.61	1.36	3.62	2.16	1.60
	(2.01)	(1.91)	(1.72)	(3.76)	(2.73)	(2.45)
Bandwidth	11.0	12	15	6.9	12	15
Observations	1617	1783	2265	689	1278	1578
	Push-Up Test			Curl Up Test		
%ED > Cutoff	-0.56	0.47	1.07	0.01	-0.59	0.03
	(2.10)	(1.54)	(1.40)	(1.93)	(1.32)	(1.21)
Bandwidth	6.5	12	15	5.8	12	15
Observations	1369	2560	3193	1302	2753	3443
	Sit and Reach			Shoulder Stretch		
%ED > Cutoff	-1.64	-1.10	-0.44	0.96	0.54	0.11
	(3.34)	(2.87)	(2.59)	(1.31)	(1.20)	(1.08)
Bandwidth	9.8	12	15	10.2	12	15
Observations	1120	1419	1780	1540	1811	2265

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. School-by-grade data from the FITNESSGRAM© test for school years spanning 2007-2011 is from the Texas Education Agency. Each coefficient is generated by a separate regression of Equation 1 using the listed fitness outcome as the dependent variable, controlling for year fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction. The sample includes all female Texas students in grades 6, 7, or 8.

Table A.9: Effects of Texas Fitness Now on Academic Outcomes, by Subgroup

	All	Female	Economically Disadvantaged
	(1)	(2)	(3)
Panel A. Pass Math TAKS			
%ED > Cutoff	0.004	0.004	0.003
	(0.010)	(0.011)	(0.011)
Observations	1,289,442	627,388	890,987
Panel B. Pass Reading TAKS			
%ED > Cutoff	0.003	0.003	0.004
	(0.006)	(0.006)	(0.007)
Observations	1,289,364	627,394	891,311
Panel C. Total Disciplinary Incidents			
%ED > Cutoff	0.149*	0.106	0.178*
	(0.090)	(0.069)	(0.103)
Observations	534,882	631,916	918,294
Panel D. Proportion of Students Disciplined			
%ED > Cutoff	0.0214*	0.0170	0.0241*
	(0.013)	(0.012)	(0.016)
Observations	1,299,744	631,916	918,294
Panel E. Number of Days Suspended			
%ED > Cutoff	0.836*	0.512**	1.007*
	(0.451)	(0.257)	(0.516)
Observations	1,299,744	631,916	918,294
Panel F. Attendance Rate			
%ED > Cutoff	-0.003**	-0.003**	-0.004***
	(0.001)	(0.001)	(0.002))
Observations	1,297,023	630,652	916,543
Bandwidth	15	15	15

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Data on test scores, disciplinary action, and attendance rates for 6th, 7th, and 8th graders is from the Education Research Center at UT-Austin for school years spanning 2007-2011. Each coefficient is generated by a separate regression of Equation 1 using the listed outcome as the dependent variable, controlling for year and grade fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction.

Table A.10: Effects of Texas Fitness Now on Academic Outcomes, by Grade

	All Grades	6th Grade	7th Grade	8th Grade
	(1)	(2)	(3)	(4)
Panel A. Pass Math TAKS				
%ED > Cutoff	0.004	0.006	0.002	0.002
	(0.010)	(0.012)	(0.012)	(0.012)
Observations	1,289,442	431,679	433,930	423,833
Panel B. Pass Reading TAKS				
%ED > Cutoff	0.003	0.000	0.006	0.001
	(0.0062)	(0.007)	(0.007)	(0.006)
Observations	1,289,364	431,677	433,926	423,761
Panel C. Total Disciplinary Incidents				
%ED > Cutoff	0.149*	0.113	0.128	0.209*
	(0.090)	(0.075)	(0.110)	(0.118)
Observations	1,299,744	433,046	435,958	430,760
Panel D. Proportion of Students Disciplined				
%ED > Cutoff	0.021*	0.017	0.020	0.027
	(0.013)	(0.012)	(0.016)	(0.015)
Observations	1,299,744	433,046	435,938	430,760
Panel E. Number of Days Suspended				
%ED > Cutoff	0.836*	0.440	0.897	1.202*
	(0.451)	(0.280)	(0.548)	(0.615)
Observations	1,299,744	433,046	435,938	430,760
Panel F. Attendance Rate				
%ED > Cutoff	-0.003**	-0.002**	-0.003**	-0.004**
	(0.001)	(0.001)	(0.002)	(0.002)
Observations	1,297,023	432,117	435,077	429,829
Bandwidth	15	15	15	15

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Data on test scores, disciplinary action, and attendance rates for 6th, 7th, and 8th graders is from the Education Research Center at UT-Austin for school years spanning 2007-2011. Each coefficient is generated by a separate regression of Equation 1 using the listed outcome as the dependent variable, controlling for year fixed effects. Standard errors are clustered on the running variable and are reported in parentheses. "%ED" represents the percent of economically disadvantaged students in the year prior to program introduction.

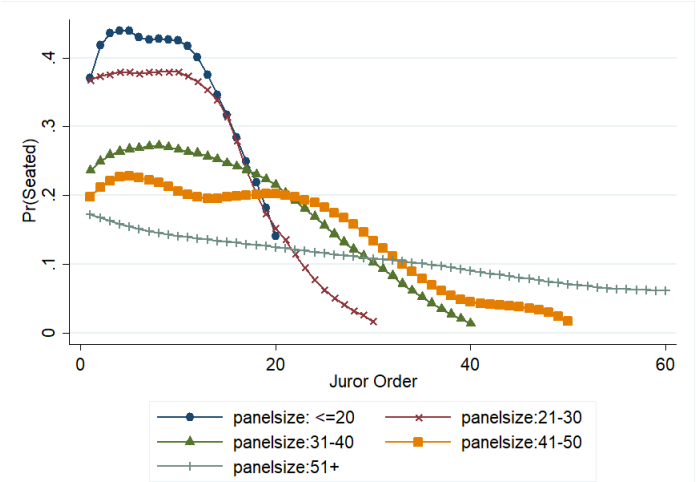
APPENDIX B

FIGURES AND TABLES FOR SECTION 3

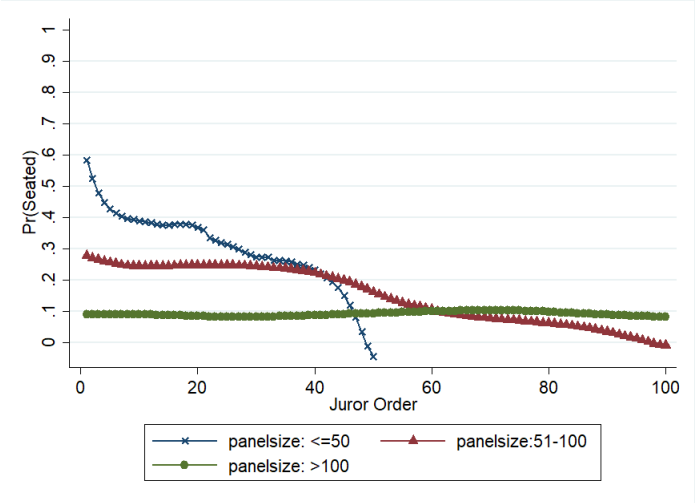
B.1 Figures

Figure B.1: Probability Seated

(a) Jury Trials with 6 Jurors



(b) Jury Trials with 12 Jurors



Notes: Each line is fit with a local linear polynomial at each panelist position using an epanechnikov kernel with varying Rule-of-Thumb (ROT) bandwidths. Figure 1a from smallest to largest panel size uses a one-sided bandwidth of 1,1,2,2, and 10. Figure 1b from smallest to largest panel size uses a one-sided bandwidth of 4, 6, and 14.

Figure B.2: Correlation between Actual Jury Gender Composition and Expected Gender Composition

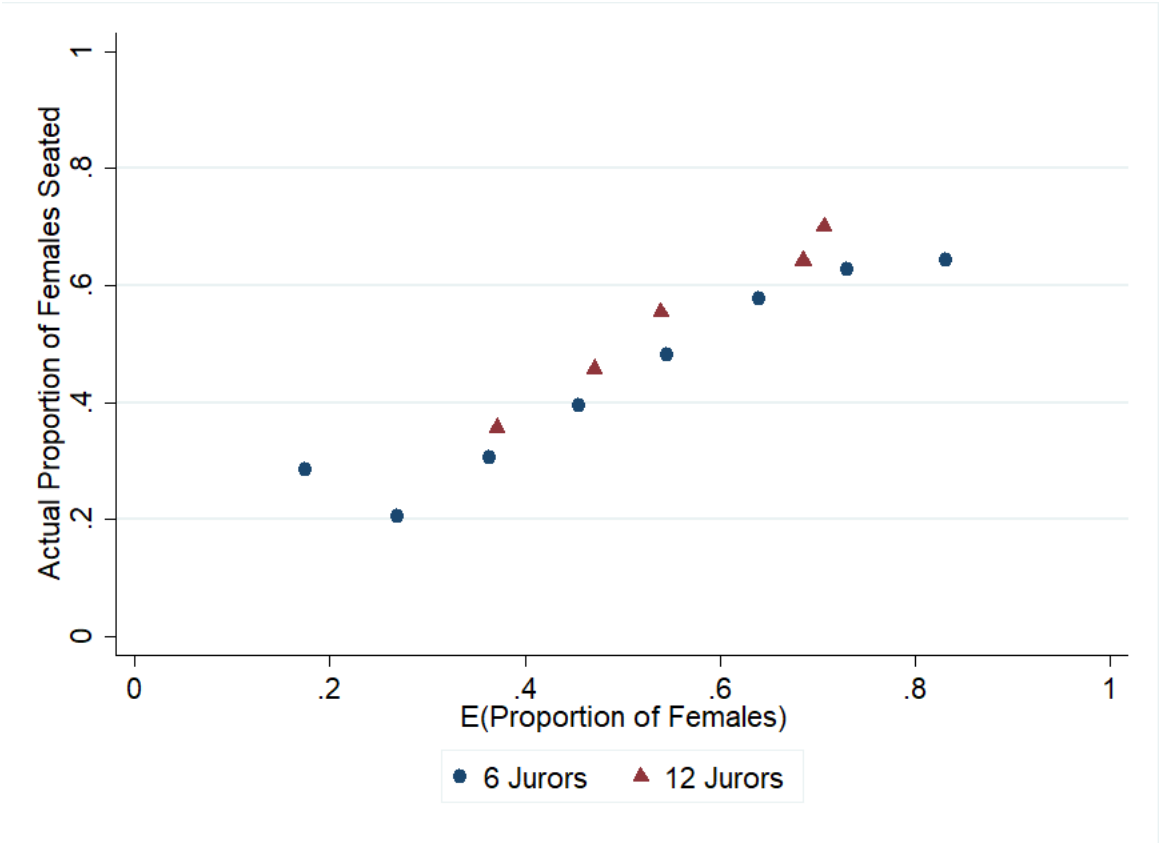
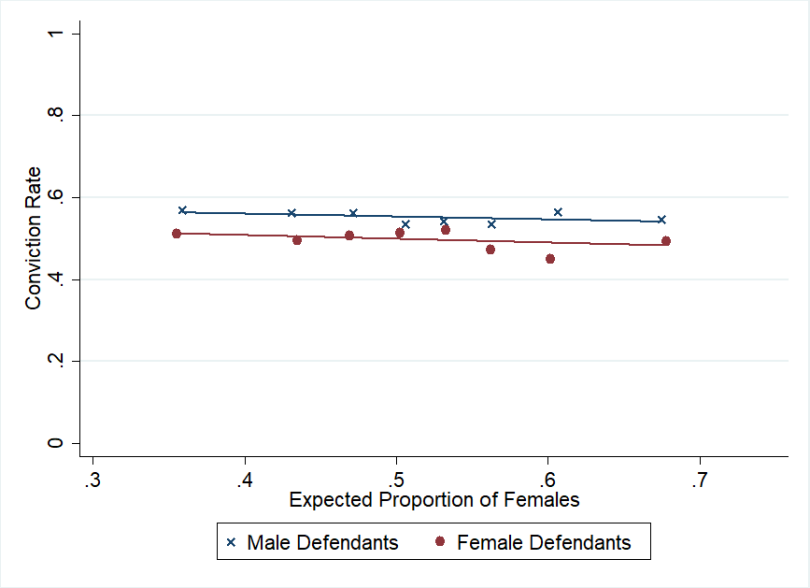
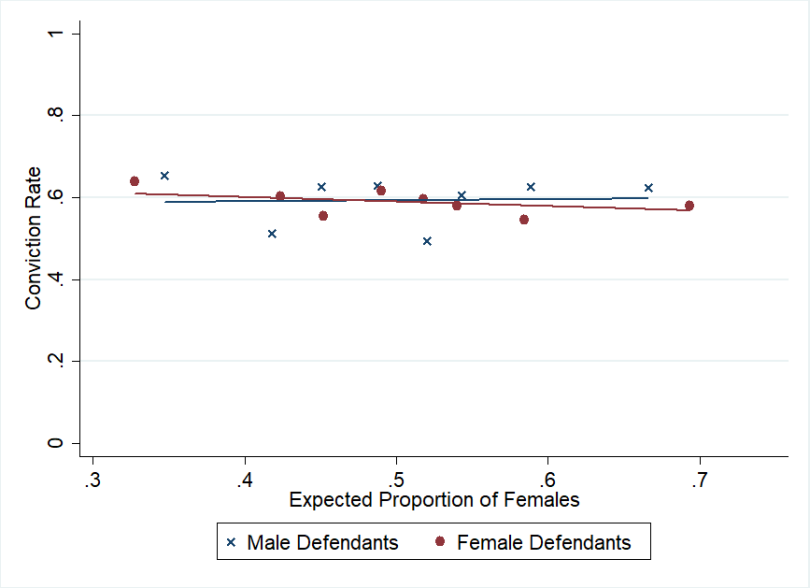


Figure B.3: Predicted Conviction Rates for Male and Female Defendants

(a) All Charges



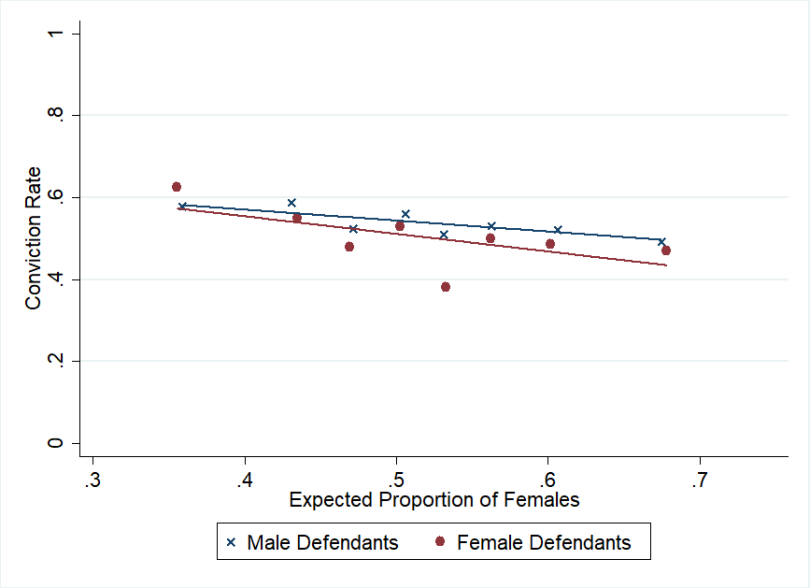
(b) Drug Charges Only



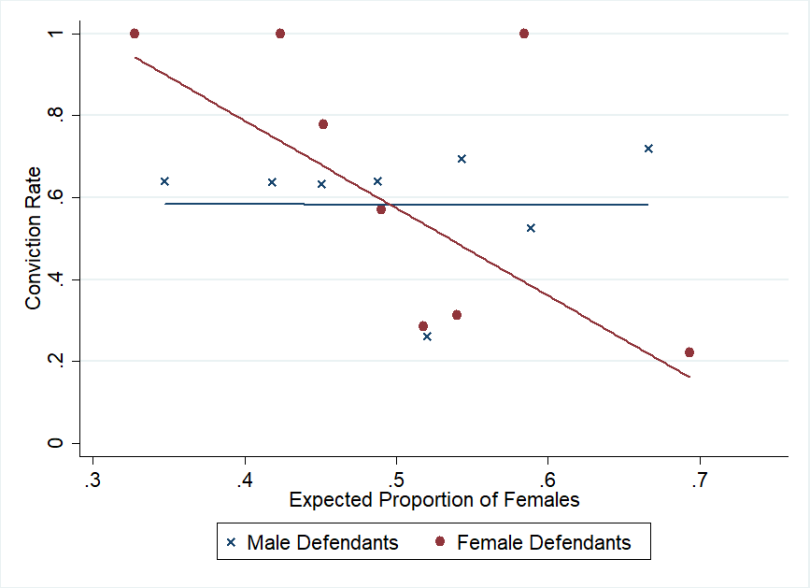
Notes: For each charge, we predict the probability of conviction using all observable characteristics. The line represents a linear fit across all predicted conviction rates.

Figure B.4: Actual Conviction Rates for Male and Female Defendants

(a) All Charges



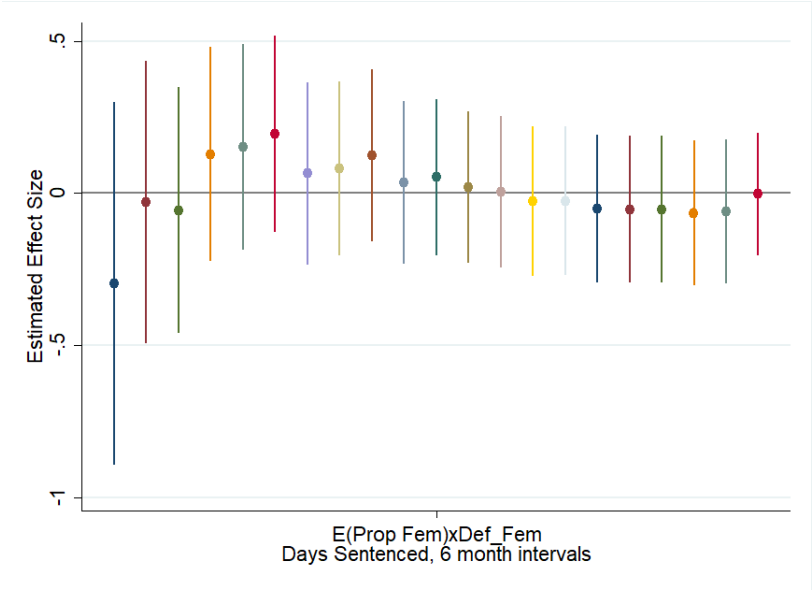
(b) Drug Charges Only



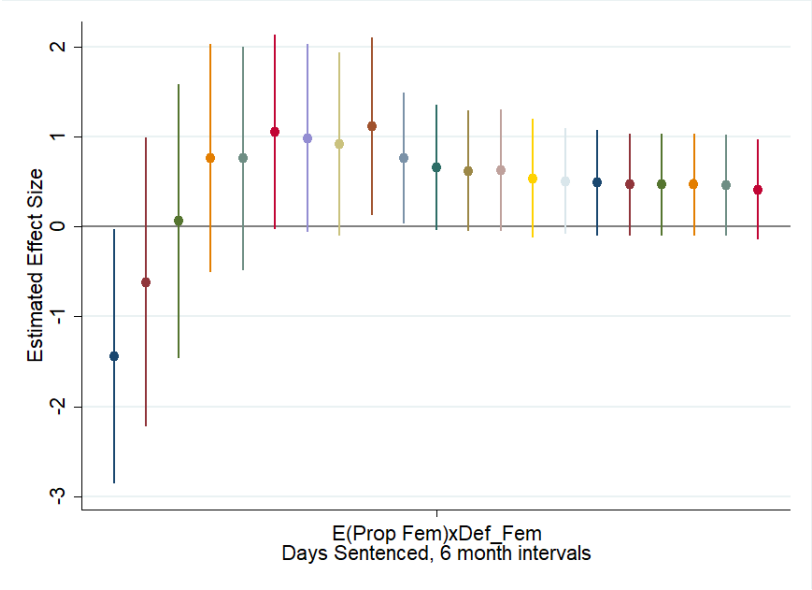
Notes: Each figure graphs the actual conviction rates for male and female defendants against the expected gender composition of the jury. Observations are grouped such that each circle represents an equal number of charges.

Figure B.5: Estimated Effects of Own-Gender Juries on Sentencing

(a) All Cases



(b) Drug-Related Cases Only



Notes: Each estimate shown represents the effect of own-gender juries on total sentencing in the case. The outcomes of interest, from left to right, are a set of indicators for sentenced to at least one day, sentenced to at least six months, 1 year, 1.5 years, 2 years, etc., up to at least 10 years. Figure 4a includes all drug, driving, property, and violent crime cases. Figure 4b restricts to cases with at least one drug charge.

B.2 Tables

Table B.1: Summary Statistics

Panel A: By Case

	All	Male	Female	Felony	Misdem.	Driving	Property	Violent	Drug
Outcomes									
Conviction Rate	0.67	0.67	0.63	0.68	0.65	0.75	0.75	0.60	0.75
Total days sentenced	1673	1931	268	2376	105	451	2251	2772	841
	(5373)	(5789)	(1126)	(6330)	(595)	(3158)	(5495)	(7440)	(1780)
P(sentenced $\geq$ 1 days)	0.42	0.45	0.27	0.50	0.26	0.34	0.54	0.43	0.50
P(sentenced $\geq$ 1 years)	0.27	0.30	0.09	0.37	0.04	0.10	0.40	0.31	0.35
P(sentenced $\geq$ 5 years)	0.16	0.19	0.04	0.23	0.01	0.04	0.25	0.23	0.13
Case Characteristics									
Defendant female	0.16	0.00	1.00	0.13	0.21	0.19	0.17	0.13	0.14
Defendant white	0.48	0.47	0.56	0.43	0.59	0.68	0.43	0.41	0.38
Defendant age	36.86	37.13	35.36	36.55	37.55	37.76	36.40	35.94	37.41
	(12.52)	(12.77)	(10.96)	(12.66)	(12.18)	(12.09)	(13.46)	(12.70)	(11.29)
Number of Charges	2.36	2.45	1.85	2.51	2.01	2.43	2.95	2.31	2.84
	(2.14)	(2.26)	(1.09)	(2.26)	(1.78)	(1.93)	(2.94)	(1.93)	(2.67)
Violent charge in case	0.47	0.49	0.38	0.57	0.25	0.07	0.27	1.00	0.10
Felony charge in case	0.69	0.71	0.57	1.00	0.00	0.37	0.84	0.83	0.74
Judge female	0.33	0.32	0.38	0.37	0.25	0.29	0.30	0.38	0.35
Jury Characteristics									
Actual Prop Female	0.46	0.45	0.48	0.45	0.47	0.47	0.45	0.45	0.44
	(0.24)	(0.24)	(0.24)	(0.25)	(0.23)	(0.23)	(0.24)	(0.24)	(0.26)
E(Proportion Female)	0.51	0.51	0.52	0.51	0.52	0.52	0.52	0.51	0.51
	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)	(0.11)
Predicted Average Juror Age	45.00	45.09	44.45	45.02	44.91	44.56	45.20	45.05	44.82
	(3.49)	(3.50)	(3.42)	(3.56)	(3.22)	(3.38)	(3.22)	(3.55)	(3.59)
Observations	1542	1302	240	1063	479	414	377	711	249

Panel B: By Charges

Outcomes									
Conviction Rate	0.53	0.54	0.50	0.56	0.47	0.52	0.57	0.50	0.58
Case Characteristics									
Defendant female	0.13	0.00	1.00	0.11	0.19	0.18	0.11	0.12	0.13
Defendant white	0.51	0.49	0.60	0.44	0.67	0.71	0.46	0.40	0.48
Defendant age	37.38	37.50	36.58	36.69	39.01	37.97	37.64	35.41	40.41
	(13.36)	(13.70)	(10.83)	(13.53)	(12.80)	(12.31)	(15.13)	(12.26)	(13.75)
Number of Charges	4.22	4.49	2.45	4.47	3.62	3.26	5.78	3.30	5.34
	(4.88)	(5.16)	(1.37)	(4.91)	(4.75)	(2.42)	(6.45)	(3.59)	(6.42)
Violent charge in case	0.43	0.43	0.37	0.54	0.16	0.09	0.18	1.00	0.09
Felony charge in case	0.70	0.72	0.57	1.00	0.00	0.29	0.90	0.87	0.72
Judge female	0.32	0.31	0.40	0.35	0.28	0.30	0.28	0.37	0.33
Jury Characteristics									
Actual Prop Female	0.46	0.46	0.45	0.45	0.48	0.47	0.48	0.44	0.45
	(0.24)	(0.24)	(0.24)	(0.25)	(0.22)	(0.21)	(0.24)	(0.24)	(0.29)
E(Proportion Female)	0.52	0.52	0.52	0.52	0.52	0.52	0.53	0.51	0.50
	(0.10)	(0.09)	(0.10)	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)	(0.10)
Predicted Average Juror Age	44.85	44.92	44.42	44.86	44.81	44.55	44.69	45.05	44.78
	(3.46)	(3.46)	(3.42)	(3.49)	(3.23)	(3.27)	(3.60)	(3.36)	(3.55)
Observations	3055	2647	408	2152	903	789	740	1056	479

Table B.2: Correlation between Actual Jury Gender Composition and Expected Gender Composition

	All	Felony	Misdemeanor	Driving	Property	Violent	Drug
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
E(Prop Fem)	0.947***	0.901***	1.040***	0.857***	0.857***	1.008***	0.919***
	(0.058)	(0.074)	(0.085)	(0.113)	(0.141)	(0.080)	(0.153)
Observations	1542	1063	451	414	377	711	249
F stat	33	17	21	35	29	80	25

Notes: Each column represents a separate regression. Columns 2 - 4 restrict the sample to cases with at least one charge in that category. All regressions include county fixed effects and columns 1-3 include county-by-crime fixed effects. Robust standard errors are in parentheses.
*p<0.10, **p<0.05, ***p<0.01

Table B.3: Exogeneity Tests

Panel A: Trial-Level								Case has at least one charge that is classified as:				
	female	white	age	avg juror age	panel size	judge female	number charges	felony	driving	property	violent	drug
E(Prop Fem)	0.082	0.177	-2.249	-1.114	-0.640	-0.043	-0.018	-0.019	-0.016	0.217**	-0.043	-0.086
	(0.098)	(0.122)	(3.228)	(1.071)	(3.268)	(0.117)	(0.387)	(0.095)	(0.112)	(0.103)	(0.127)	(0.096)
Observations	1542	1542	1542	839	1542	1542	1542	1542	1542	1542	1542	1542
Panel B: Charge-Level												
	female	white	age	avg juror age	panel size	judge female	number charges	felony	driving	property	violent	drug
E(Prop Fem)	0.075	0.175	-2.025	-0.936	-1.205	-0.027	0.075	-0.053	-0.015	0.164*	-0.039	-0.094
	(0.099)	(0.123)	(3.266)	(1.079)	(3.237)	(0.118)	(0.412)	(0.096)	(0.106)	(0.094)	(0.124)	(0.086)
Observations	3056	3056	3056	1498	3056	3056	3056	3056	3056	3056	3056	3056

Notes: Each column in each panel reports estimates from a separate regression in which we regress observable characteristics on the expected proportion of females on the jury. Columns 1 - 7 include county-by-crime fixed effects, and columns 8 - 12 include county fixed effects. The first three columns show results for defendant characteristics. Standard errors are in parentheses and are clustered at the defendant level.

*p<0.10, **p<0.05, ***p<0.01

Table B.4: Effect of Own-Gender Juries on Conviction Rates, by Severity

	All Charges			Felony Charges			Misdemeanor Charges		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
E(Prop Fem)xDef_Fem	-0.247	-0.247	-0.328	-0.323	-0.327	-0.468	-0.474	-0.483	-0.470
	(0.306)	(0.307)	(0.309)	(0.418)	(0.419)	(0.408)	(0.410)	(0.411)	(0.421)
	(0.42049)	(0.42001)	(0.28894)	(0.43976)	(0.43495)	(0.25230)	(0.24789)	(0.23991)	(0.26461)
Observations	3056	3056	3056	1726	1726	1726	1330	1330	1330
Mean Dependent Variable	0.54	0.54	0.54	0.54	0.54	0.54	0.52	0.52	0.52
Def & Jury Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CountyXCrime Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Interactions	No	No	Yes	No	No	Yes	No	No	Yes

Notes:All specifications include controls for defendant gender and expected gender composition of the jury, as well as county-by-crime fixed effects. Additional controls include defendant age, the number of charges in the case, and indicators for defendant's race, judge's gender, and whether there was charge for a violent crime in the case. Interactions include controls for each of those characteristics interacted with the expected proportion of female jurors. Standard errors are in parentheses and are clustered at the defendant level. False discovery rate (FDR) adjusted Q-values adjust for multiple inference given the six subcategories of crime examined. They are constructed using the method proposed by Anderson (2008) and are interpreted as two-sided p-values.
*p<0.10, **p<0.05, ***p<0.01

Table B.5: Effect of Own-Gender Juries on Conviction Rates, by Crime Type

	Driving Charges			Property Charges			Violent Charges			Drug Charges		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
E(Prop Fem)xDef_Fem	0.154	0.118	0.070	0.610	0.540	0.543	-0.392	-0.297	-0.237	-2.194***	-2.185***	-1.804**
	(0.510)	(0.505)	(0.540)	(0.666)	(0.677)	(0.635)	(0.498)	(0.513)	(0.505)	(0.596)	(0.596)	(0.723)
	(0.763167)	(0.815713)	(0.897007)	(0.360900)	(0.425471)	(0.393430)	(0.431619)	(0.562634)	(0.638804)	(0.000290)	(0.000304)	(0.013337)
Observations	789	789	789	740	740	740	1057	1057	1057	479	479	479
Mean Dependant Variable	0.55	0.55	0.55	0.53	0.53	0.53	0.50	0.50	0.50	0.64	0.64	0.64
Def & Jury Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Interactions	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

Notes: All specifications include controls for defendant gender and expected gender composition of the jury, as well as county fixed effects. Additional controls include defendant age, the number of charges in the case, and indicators for defendant's race, judge's gender, and whether there was charge for a violent crime in the case. Interactions include controls for each of those characteristics interacted with the expected proportion of female jurors.

Standard errors are in parentheses and are clustered at the defendant level. False discovery rate (FDR) adjusted Q-values adjust for multiple inference given the six subcategories of crime examined. They are constructed using the method proposed by Anderson (2008) and are interpreted as two-sided p-values.

*p<0.10, **p<0.05, ***p<0.01

Table B.6: Effect of Own-Gender Juries on Being Sentenced to Jail

	All Charges			Drug Charges		
	(1)	(2)	(3)	(4)	(5)	(6)
E(Prop Fem)xDef_Fem	-0.296	-0.220	-0.284	-1.453**	-1.432*	-1.264*
	(0.304)	(0.296)	(0.299)	(0.719)	(0.756)	(0.704)
	(0.3307)	(0.4578)	(0.3411)	(0.0445)	(0.0596)	(0.0740)
Observations	1534	1534	1534	245	245	245
Mean Dependant Variable	0.41	0.41	0.41	0.49	0.49	0.49
Def & Jury Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	No	Yes	Yes
Interactions	No	No	Yes	No	No	Yes

Notes: All specifications include controls for defendant gender and expected gender composition of the jury, as well as county fixed effects. Additional controls include defendant age, the number of charges in the case, and indicators for defendant's race, judge's gender, and whether there was charge for a violent crime in the case. Interactions include controls for each of those characteristics interacted with the expected proportion of female jurors and defendant's gender.

Standard errors are in parentheses.

*p<0.10, **p<0.05, ***p<0.01

Table B.7: Robustness of Estimates of Own-Gender Juries on Conviction Rates - Drug Charges Only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
E(Prop Fem)XDef_Fem	-1.815**	-2.092***	-2.296**	-1.659**	-1.614**	-1.844**	-1.749**	-1.681**	-1.887***	-1.698**
	(0.724)	(0.772)	(0.889)	(0.743)	(0.639)	(0.757)	(0.698)	(0.718)	(0.723)	(0.693)
Observations	479	479	295	479	479	479	479	479	479	479
Mean Dependant Variable	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64
Def & Jury Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Interactions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Drug Type Interactions	No	Yes	No	No	No	No	No	No	No	No
Juror Age Control & Interaction	No	No	Yes	No	No	No	No	No	No	No
Adjudication Withheld=Not Guilty	No	No	No	Yes	No	No	No	No	No	No
Missing Genders	half	half	half	half	female	male	half	half	half	half
Predicted Genders	API	API	API	API	API	API	SS	API	API	API
Pr(Seated)	LL	LL	LL	LL	LL	LL	LL	Raw	Probit	LL
Pr(SeatedlPanelsize)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

Notes: Standard errors are in parentheses and are clustered at the defendant level.
*p<0.10, **p<0.05, ***p<0.01

B.3 Additional Results

Table B.8: Effect of Own-Gender Juries on Conviction Rates

	All	Felony	Misdemeanor	Drug	Driving	Property	Violent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
E(Prop Fem)	-0.110	-0.131	-0.018	0.130	-0.425*	-0.049	0.002
	(0.114)	(0.146)	(0.159)	(0.195)	(0.243)	(0.181)	(0.290)
Observations	3055	1725	1330	789	740	1056	479

Notes: Each column represents a separate regression. Columns 1 - 3 include county-by-crime fixed effects, and columns 4 - 7 include county fixed effects. Standard errors are in parentheses and are clustered at the defendant level. *p<0.10, **p<0.05, ***p<0.01

Table B.9: Exogeneity Tests with Actual Proportion of Female Jurors

Panel A: Trial-Level								Case has at least one charge that is classified as:				
	female	white	age	avg juror age	panel size	judge female	number charges	felony	driving	property	violent	drug
Actual Proportion Female	0.090*	0.137**	-0.536	-1.089*	8.672	-0.007	0.081	-0.007	0.002	0.102*	-0.120*	0.010
	(0.050)	(0.068)	(1.865)	(0.606)	(6.585)	(0.065)	(0.301)	(0.051)	(0.060)	(0.056)	(0.069)	(0.052)
Observations	1542	1542	1542	839	1542	1542	1542	1542	1542	1542	1542	1542
Panel B: Charge-Level												
	female	white	age	avg juror age	panel size	judge female	number charges	felony	driving	property	violent	drug
Actual Prop Female	0.090*	0.138**	-0.265	-1.071*	8.530	-0.009	0.068	0.002	-0.003	0.113**	-0.100	-0.007
	(0.050)	(0.068)	(1.877)	(0.603)	(6.674)	(0.065)	(0.302)	(0.051)	(0.059)	(0.050)	(0.068)	(0.047)
Observations	3055	3055	3055	1497	3055	3055	3055	3055	3055	3055	3055	3055

Notes: Each column in each panel reports estimates from a separate regression in which we regress observable characteristics on the actual proportion of females on the seated jury. Columns 1 - 7 include county-by-crime fixed effects, and columns 8 - 12 include county fixed effects. The first three columns show results for defendant characteristics. Standard errors are in parentheses and are clustered at the defendant level. Standard errors are in parentheses and are clustered at the defendant level.
*p<0.10, **p<0.05, ***p<0.01

Table B.10: Effect of Own-Gender Juries on Conviction Rates, by Jury Trial Status

	Non-Trial			Jury Trial		
	(1)	(2)	(3)	(4)	(5)	(6)
E(Prop Fem)xDef_Fem	-2.419***	-2.522***	-2.960**	-5.760***	-6.416***	-5.637**
	(0.663)	(0.717)	(1.393)	(1.309)	(1.314)	(2.491)
Observations	177	177	177	165	165	165
Mean Dependant Variable	0.64	0.64	0.64	0.64	0.64	0.64
Def & Jury Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	No	Yes	Yes
Interactions	No	No	Yes	No	No	Yes

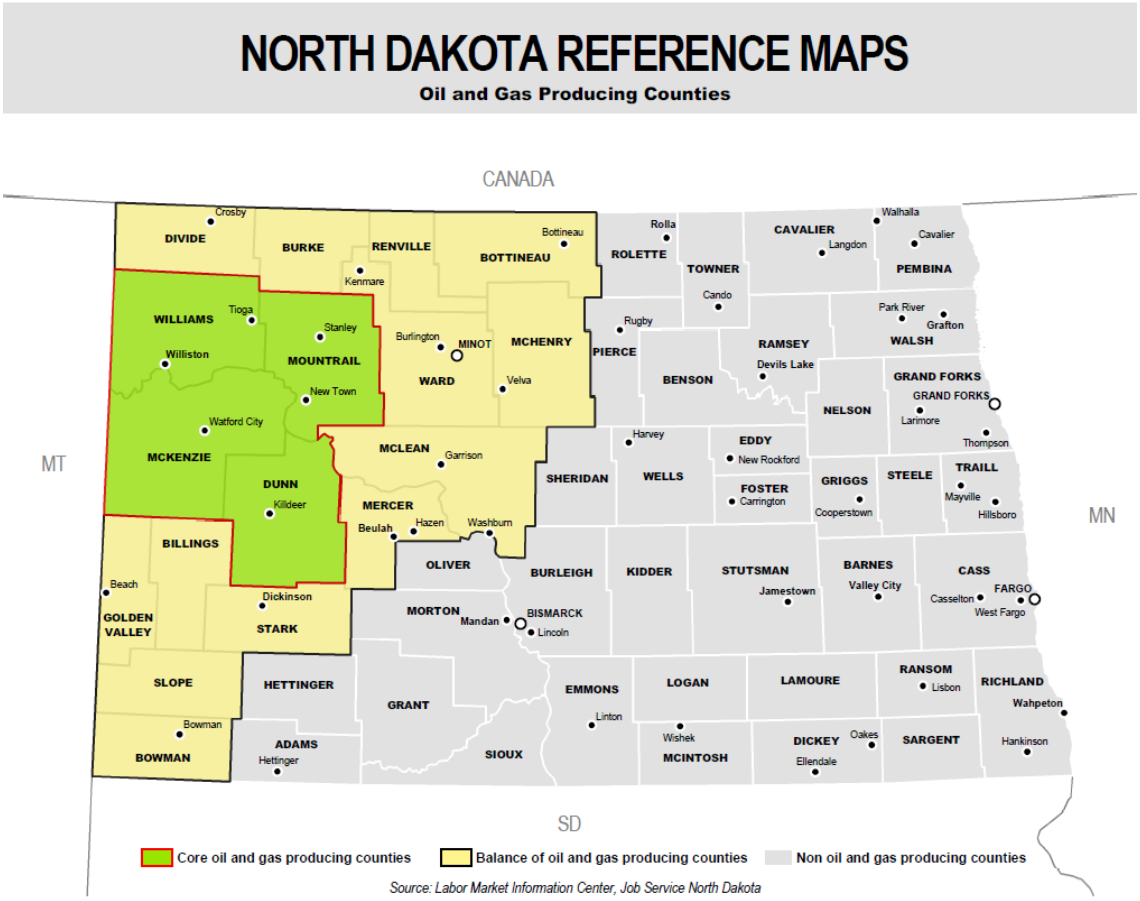
Notes: All specifications include controls for defendant gender and expected gender composition of the jury, as well as county fixed effects. Additional controls include defendant age, the number of charges in the case, and indicators for defendant's race, judge's gender, and whether there was charge for a violent crime in the case. Interactions include controls for each of those characteristics interacted with the expected proportion of female jurors.
Standard errors are in parentheses and are clustered at the defendant level.
*p<0.10, **p<0.05, ***p<0.01

APPENDIX C

FIGURES AND TABLES FOR SECTION 4

C.1 Figures

Figure C.1: Fracking Counties reprinted from North Dakota Labor Market Information (2018)

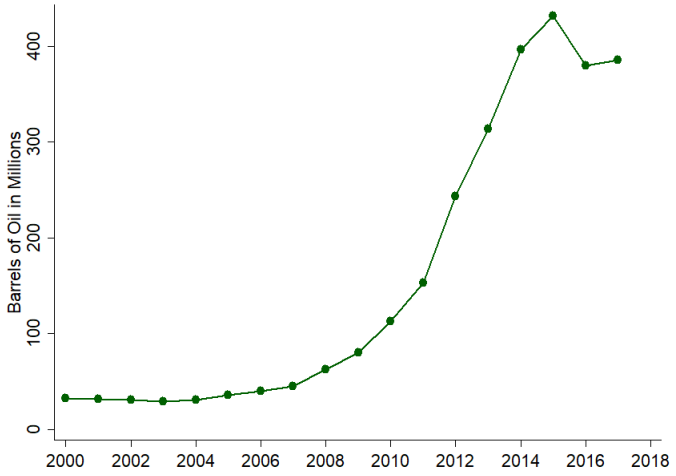


Source: Labor Market Information, Job Service North Dakota

Figure C.2: Leasing and Production

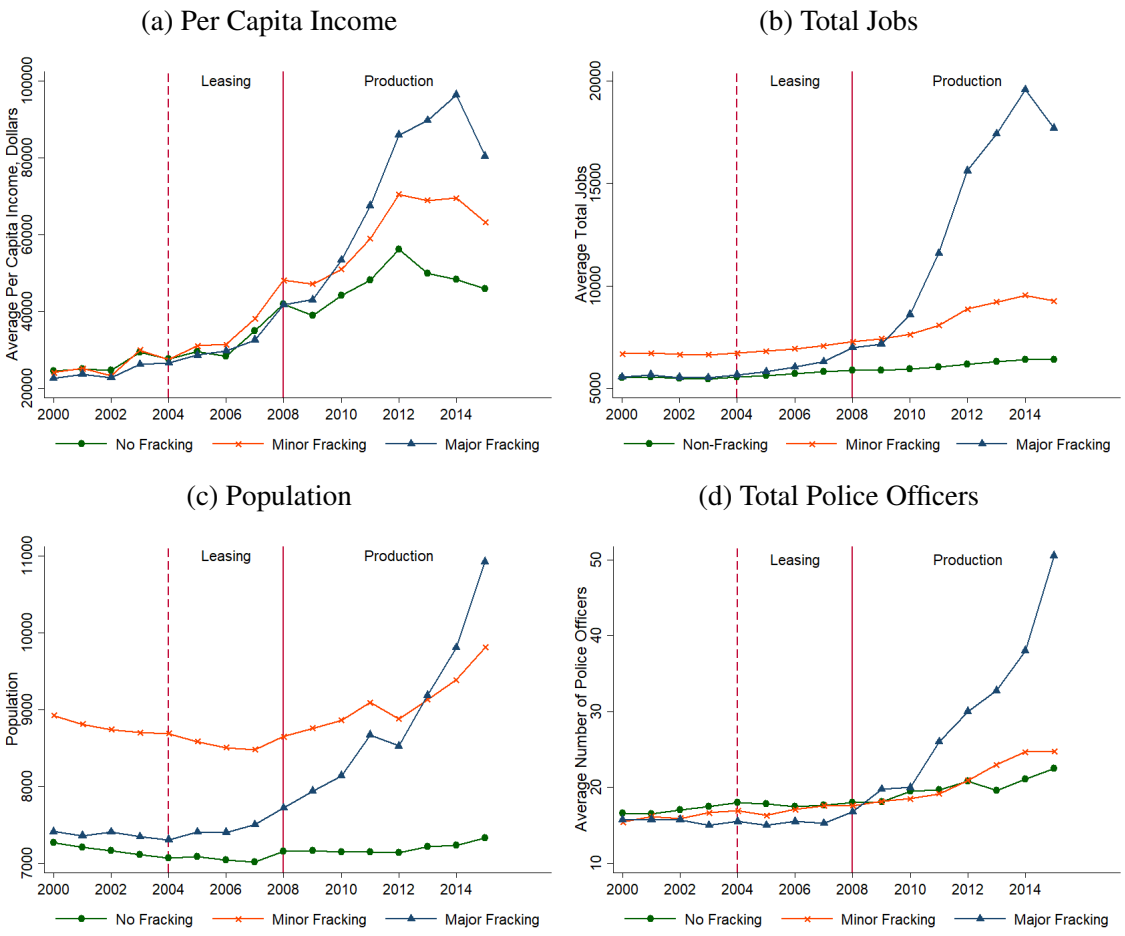
(a) Leases

(b) Production



Notes: All leases in North Dakota are collected from Drilling Info for 2000-2017. Only leases matched to rural residents in the early 2000s are depicted in the figure above, as this is the sample of leases used in the analysis. Monthly county production data are from North Dakota Department of Mineral Resources.

Figure C.3: County Demographics by Fracking Region

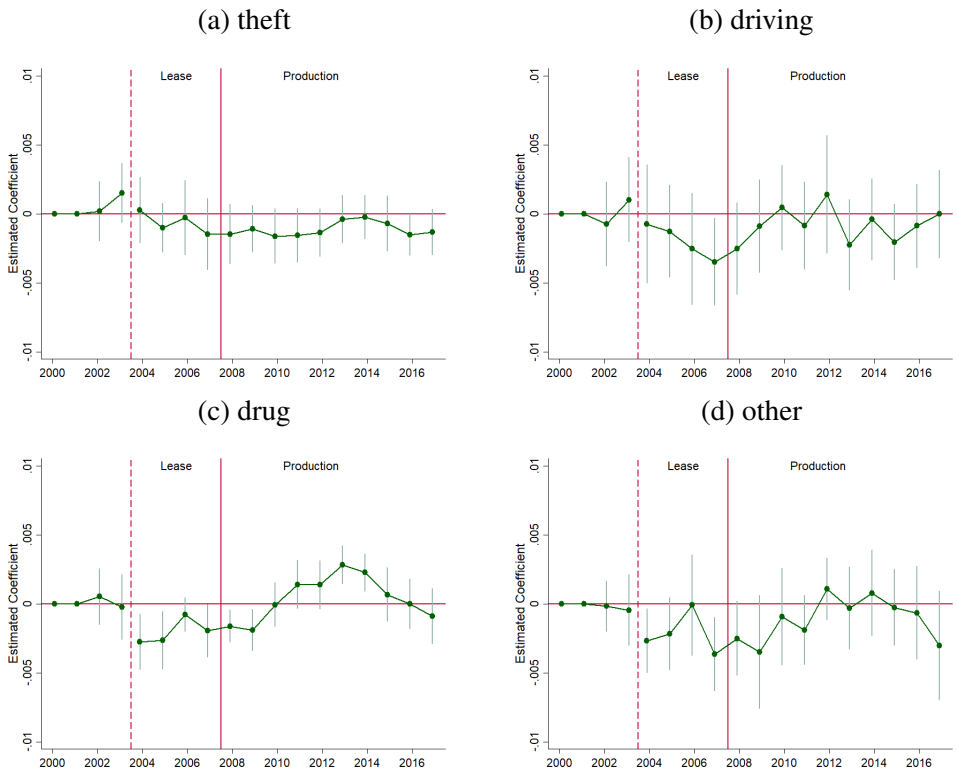


Notes: Data on income and jobs are from Bureau of Economic Analysis. Population is calculated using the number of migrant and non-migrant tax exemptions from the Internal Revenue Service. Police employment data are from the Uniform Crime Reporting Program: Police Employee (LEOKA) Data.

Figure C.4: Dynamic Difference-in-Difference Estimates of the Effect of Fracking on Crime

Notes: Dynamic difference-in-differences estimates from equation 1. Standard errors are clustered at the county-level and 95% confidence intervals are shown. Data are from the State of North Dakota Judicial Branch from 2000-2017.

Figure C.5: Dynamic Difference-in-Difference Estimates of the Effect of Fracking on Crime, by Crime Type



Notes: Dynamic difference-in-differences estimates from equation 1 with household and year fixed effects. Standard errors are clustered at the county-level and 95% confidence intervals are shown. Data are from the State of North Dakota Judicial Branch from 2000-2017.

Figure C.6: Dynamic Difference-in-Difference Estimates of the Effect of Fracking on Police and Population

(a) Population

(b) Total Officers

Notes: Dynamic difference-in-differences estimates from equation 1 at the county-level. Standard errors are clustered at the county-level and 95% confidence intervals are shown. Data are from Internal Revenue Service and Uniform Crime Reporting Program Data [United States]: Police Employee (LEOKA).

Figure C.7: Estimates of the Effect of Fracking on Out-Migration

Notes: Dynamic difference-in-differences estimates from Equation 1 with county and year fixed effects. Standard errors are clustered at the county-level and 95% confidence intervals are shown. Outcome is defined as total number of out-migration exemptions. An exemption is classified as a migrant if it is filed in a different county than in the the previous year. The exemption would be an out-migrant for the county of filing in the previous year and an in-migrant for the county of filing in the current year. Data on all exemptions is from the Internal Revenue Service.

C.2 Tables

Table C.1: Summary Statistics

	All	Fracking County	Non-Fracking County	Lease Holder	Non-Lease Holder
Panel A: Household					
Case ever filed	0.20	0.20	0.20	0.22	0.19
Lease holder	0.21	0.42	0.08	1.00	0.00
Monthly Payment	1141.23	1262.10	143.73	10945.82	0.00
	(9907.74)	(10325.90)	(2930.07)	(28886.64)	(0.00)
Number of months	5.29	8.04	0.83	50.78	0.00
	(22.11)	(25.94)	(8.73)	(48.77)	(0.00)
Observations	31169	6964	21394	6436	24733
Panel B: Household-Year					
Case filed	0.0232	0.0219	0.0238	0.0283	0.0218
Drug charge	0.0047	0.0035	0.0050	0.0060	0.0043
Driving charge	0.0132	0.0129	0.0134	0.0166	0.0124
Theft charge	0.0043	0.0041	0.0045	0.0048	0.0042
Other charge	0.0085	0.0081	0.0088	0.0105	0.0080
Observations	561042	125352	385092	115848	445194
Panel C: Charges					
Charges per case	1.14	1.11	1.15	1.15	1.14
	(0.60)	(0.50)	(0.64)	(0.58)	(0.61)
Felony charge	0.10	0.09	0.11	0.11	0.10
Driving charge	0.44	0.45	0.42	0.44	0.44
Drug charge	0.17	0.13	0.18	0.18	0.17
Theft charge	0.17	0.17	0.17	0.15	0.17
Assault charge	0.04	0.03	0.04	0.04	0.04
Other charge	0.30	0.30	0.30	0.31	0.30
Male	0.78	0.77	0.79	0.78	0.79
Age	34.38	33.86	34.37	33.90	34.55
	(14.30)	(14.02)	(14.35)	(13.81)	(14.47)
Observations	23091	4746	16583	6076	17015

Table C.2: Estimates of the Effect of Fracking on Crime

	1	2	3	4
Fracking Co X Post Lease	-0.0044***	-0.0048***	-0.0050***	-0.0050**
	(0.0015)	(0.0017)	(0.0017)	(0.0019)
Fracking Co X Post Prod	-0.0027	-0.0031*	-0.0043*	-0.0018
	(0.0020)	(0.0019)	(0.0025)	(0.0022)
Pre Lease		-0.0009		
		(0.0014)		
Observations	561078	561078	561078	561078
Mean Dependent Variable	0.02	0.02	0.02	0.02
Household & Year FE	Y	Y	Y	Y
Lead	N	Y	N	N
County Trends	N	N	Y	N
Pre-Period County Controls X Year	N	N	N	Y

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are in parentheses and clustered at the county level. County controls include per capita income, total jobs, population, total officers, and production in 2000.

Table C.3: Estimates of the Effect of Fracking on Crime, by Crime Type

	1	2	3	4
Panel A: Property Case Filed				
Fracking Co X Post Lease	-0.0011	-0.0006	-0.0012	-0.0016**
	(0.0008)	(0.0010)	(0.0009)	(0.0007)
Adjusted FDR Q-values	[0.316]	[0.519]	[0.242]	[0.119]
Fracking Co X Post Prod	-0.0016**	-0.0011	-0.0020*	-0.0016*
	(0.0007)	(0.0007)	(0.0011)	(0.0009)
Adjusted FDR Q-values	[0.076]	[0.316]	[0.191]	[0.211]
Pre Lease		0.0008		
		(0.0008)		
Observations	561078	561078	561078	561078
Mean Dependant Variable	0.004	0.004	0.004	0.004
Panel B: Driving Case Filed				
Fracking Co X Post Lease	-0.0021*	-0.0020	-0.0020*	-0.0022
	(0.0012)	(0.0015)	(0.0012)	(0.0016)
Adjusted FDR Q-values	[0.188]	[0.374]	[0.195]	[0.292]
Fracking Co X Post Prod	-0.0009	-0.0008	-0.0008	-0.0002
	(0.0012)	(0.0012)	(0.0017)	(0.0015)
Adjusted FDR Q-values	[0.535]	[0.519]	[0.651]	[0.879]
Pre Lease		0.0001		
		(0.0013)		
Observations	561078	561078	561078	561078
Mean Dependant Variable	0.013	0.013	0.013	0.013
Panel C: Drug Case Filed				
Fracking Co X Post Lease	-0.0021**	-0.0020***	-0.0028***	-0.0024***
	(0.0009)	(0.0007)	(0.0010)	(0.0009)
Adjusted FDR Q-values	[0.076]	[0.049]	[0.062]	[0.052]
Fracking Co X Post Prod	0.0003	0.0004	-0.0015	-0.0001
	(0.0005)	(0.0005)	(0.0012)	(0.0005)
Adjusted FDR Q-values	[0.535]	[0.519]	[0.254]	[0.875]
Pre Lease		0.0001		
		(0.0008)		
Observations	561078	561078	561078	561078
Mean Dependant Variable	0.005	0.005	0.005	0.005
Panel D: Other Case Filed				
Fracking Co X Post Lease	-0.0020**	-0.0022**	-0.0025**	-0.0015
	(0.0009)	(0.0011)	(0.0010)	(0.0011)
Adjusted FDR Q-values	[0.076]	[0.192]	[0.062]	[0.292]
Fracking Co X Post Prod	-0.0010	-0.0011	-0.0022	-0.0004
	(0.0012)	(0.0012)	(0.0014)	(0.0012)
Adjusted FDR Q-values	[0.535]	[0.519]	[0.208]	[0.875]
Pre Lease		-0.0003		
		(0.0009)		
Observations	561078	561078	561078	561078
Mean Dependant Variable	0.008	0.008	0.008	0.008
Household & Year FE	Y	Y	Y	Y
Lead	N	Y	N	N
County Linear Trends	N	N	Y	N
Pre-Period County Controls X Year	N	N	N	Y

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are in parentheses and clustered at the county level. County controls include per capita income, total jobs, population, total officers, and production in 2000.

Table C.4: Estimates of the Effect of Fracking on Crime, by Intensity

	1	2	3	4
Minor Fracking County X Post Lease	-0.0039**	-0.0044***	-0.0039**	-0.0044**
	(0.0015)	(0.0016)	(0.0017)	(0.0021)
Major Fracking County X Post Lease	-0.0055*	-0.0060*	-0.0075***	-0.0070**
	(0.0030)	(0.0032)	(0.0028)	(0.0032)
Minor Fracking County X Post Prod	-0.0035*	-0.0040**	-0.0036	-0.0019
	(0.0021)	(0.0019)	(0.0030)	(0.0023)
Major Fracking County X Post Prod	-0.0007	-0.0012	-0.0062**	-0.0016
	(0.0034)	(0.0035)	(0.0026)	(0.0032)
Pre Lease		-0.0009		
		(0.0014)		
Observations	561078	561078	561078	561078
Mean Dependent Variable	0.02	0.02	0.02	0.02
Household & Year FE	Y	Y	Y	Y
Lead	N	Y	N	N
County Linear Trends	N	N	Y	N
Pre-Period County Controls X Year	N	N	N	Y

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are in parentheses and clustered at the county level. County controls include per capita income, total jobs, population, total officers, and production in 2000.

Table C.5: Estimates of the Effect of Fracking on Crime, by Lease Status

	1	2	3	4
Lease HH X Post Lease	-0.0021	-0.0026	-0.0028	-0.0023
	(0.0019)	(0.0021)	(0.0020)	(0.0023)
Lease HH X Post Prod	-0.0008	-0.0012	-0.0027	0.0007
	(0.0023)	(0.0021)	(0.0026)	(0.0026)
Non-Lease HH X Post Lease	-0.0066***	-0.0070***	-0.0070***	-0.0068***
	(0.0018)	(0.0020)	(0.0020)	(0.0020)
Non-Lease HH X Post Prod	-0.0046**	-0.0050**	-0.0059**	-0.0034
	(0.0021)	(0.0021)	(0.0027)	(0.0021)
Pre Lease		-0.0009		
		(0.0014)		
Observations	561078	561078	561078	561078
Mean Dependent Variable	0.02	0.02	0.02	0.02
Household & Year FE	Y	Y	Y	Y
Lead	N	Y	N	N
County Linear Trends	N	N	Y	N
Pre-Period County Controls X Year	N	N	N	Y

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are in parentheses and clustered at the county level. County controls include per capita income, total jobs, population, total officers, and production in 2000.

C.3 Additional Results

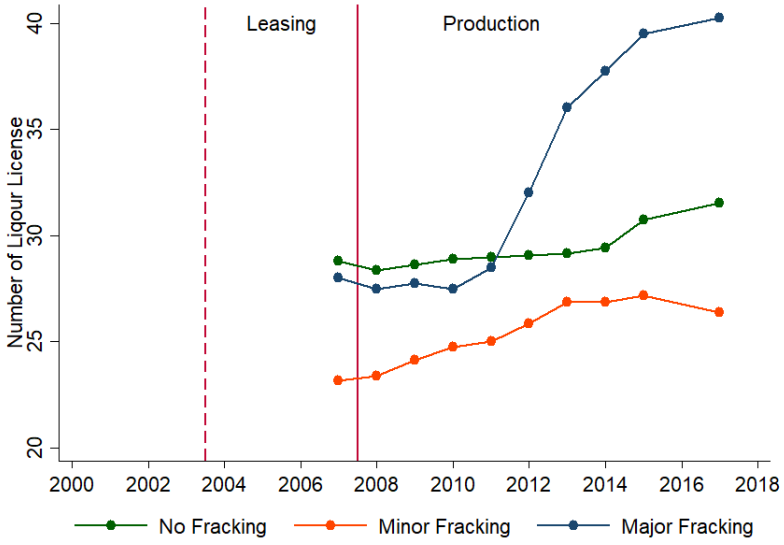
Figure C.8: Estimates of the Effect of Fracking on Real Estate

(a) Total Sales

(b) Price

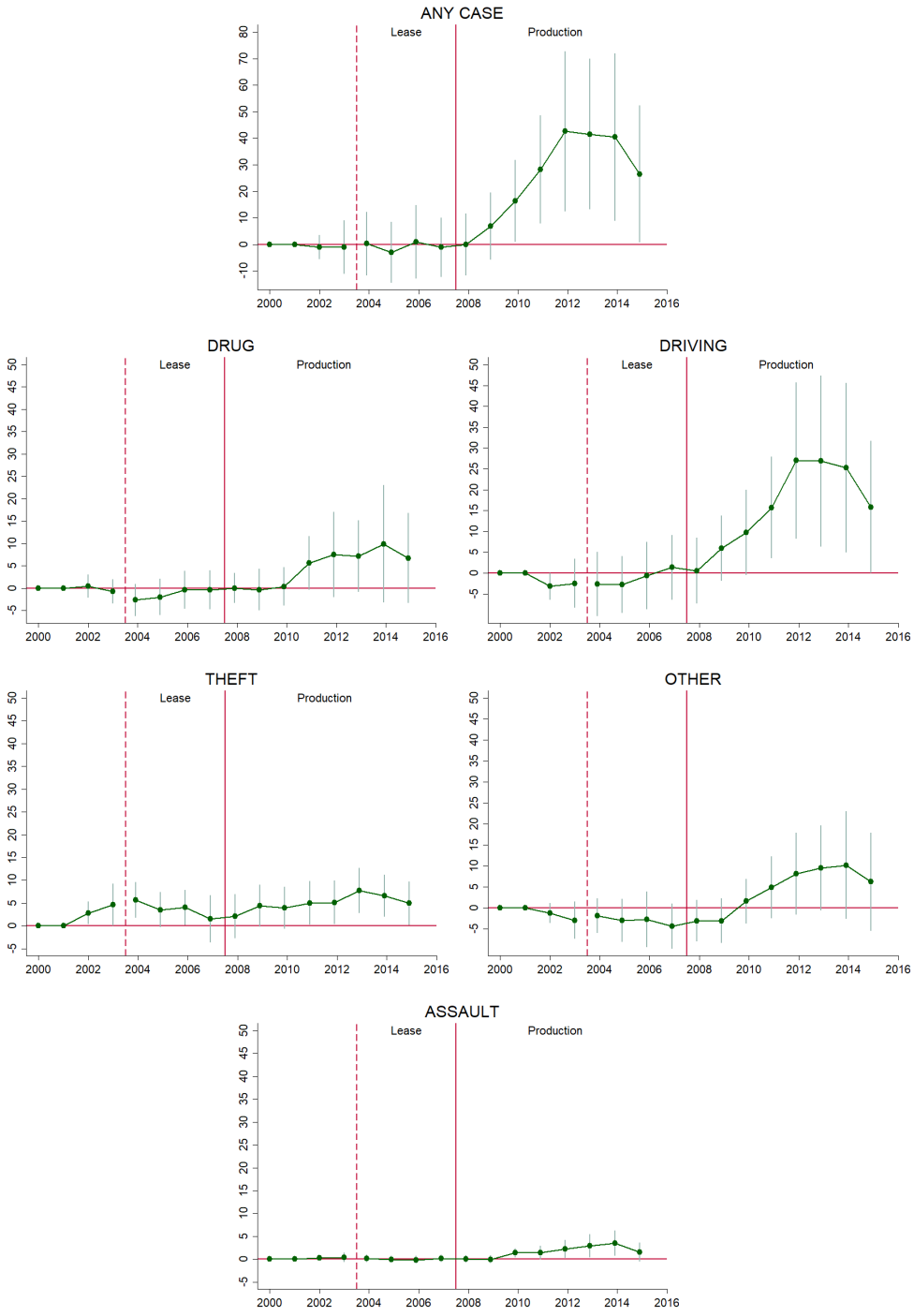
Notes: Dynamic difference-in-differences estimates from Equation 1 with county and year fixed effects. Standard errors are clustered at the county-level and 95% confidence intervals are shown. Outcome is defined as total sales in each county and total sale values. Data on all property sales are from the North Dakota State Board of Equalization.

Figure C.9: Average Total Number of Liqour Licenses per County by Fracking Region



Notes: Data on all liqour licenses in the State of North Dakota are provided by the North Dakota Attorney General's office from 2007-2018. Average total number of of licenses per county by fracking region are plotted.

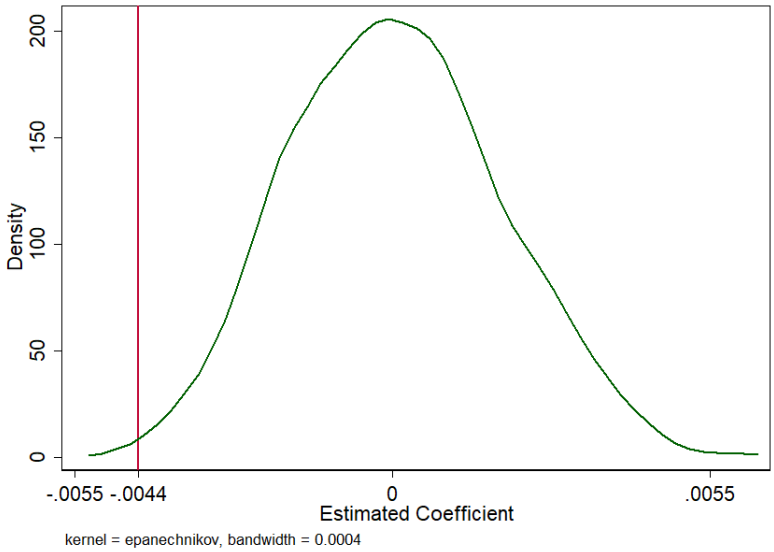
Figure C.10: Estimates of the Effect of Fracking on Aggregate Crime, Residents and Non-Residents



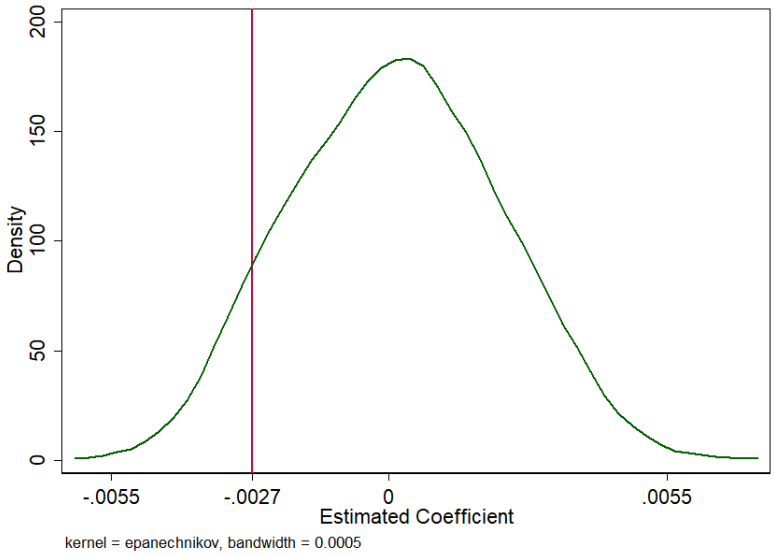
Notes: Dynamic difference-in-differences estimates from equation 1 with county and year fixed effects. Standard errors are clustered at the county-level and 95% confidence intervals are shown. Outcome is defined as cases filed per 1000 persons with population measured using IRS tax exemptions in each year.

Figure C.11: Placebo Tests

(a) Placebo Estimates for Any Case during Leasing Period



(b) Placebo Estimates for Any Case during Production Period



Notes: Figure plots the density of 1000 estimates from equation 1 with fracking status randomly assigned to 17 counties. The red line in Figure A.2a depicts the main estimate during leasing period, -0.0044, with 1 estimate less than or equal to it. Similarly, in Figure A.2b the estimate during production period, -0.0027, is drawn in red with 88 estimates less than or equal to it.

Table C.6: Case Filed, Robustness to Levenshtein Index

	1	2	3	4
Fracking Co X Post Lease	-0.0044***	-0.0044***	-0.0035**	-0.0031**
	(0.0015)	(0.0015)	(0.0013)	(0.0012)
Fracking Co X Post Production	-0.0009	-0.0027	-0.0032*	-0.0032*
	(0.0026)	(0.0020)	(0.0018)	(0.0016)
Observations	562500	561132	560592	560340
Mean Dependent Variable	0.02	0.02	0.02	0.02
Household & Year FE	Y	Y	Y	Y
Levenshtein Distance	3	2	1	0

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are in parentheses and clustered at the county level. Levenshtein Distance is the number of string edits permitted when match households using last name, street number, city, and zip code. Column 1 allows for three string edits when matching households, which is one more than what is used throughout the paper. Column 2 replicates Column 1 from Table 2 with two string edits as a baseline specification. Column 2 and 3 restrict to matches with a string distance of one or zero, respectively.

Table C.7: Estimates of the Effect of Fracking on Crime, Robust to Functional Form and Intensive Margin

Dependant Variable	Any Case		Number of Cases		Number of Charges	
	1	2	3	4	5	6
Fracking Co X Post Lease	-0.0044***	-0.2102***	-0.0059***	-0.2399***	-0.0059***	-0.1927***
	(0.0015)	(0.0781)	(0.0019)	(0.0768)	(0.0019)	(0.0671)
Fracking Co X Post Prod	-0.0027	-0.1476	-0.0036	-0.1823**	-0.0039*	-0.1517*
	(0.0020)	(0.1052)	(0.0022)	(0.0877)	(0.0022)	(0.0852)
Observations	561042	110034	561042	110052	561042	110052
Mean Dependent Variable	0.02	0.02	0.04	0.04	0.04	0.04
Household & Year FE	Y	Y	Y	Y	Y	Y
Ordinary Least Squares	Y	N	N	N	N	N
Logit	N	Y	N	N	N	N
Inverse Hyperbolic Sine	N	N	Y	N	Y	N
Poisson	N	N	N	Y	N	Y

Notes: *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. Standard errors are in parentheses and clustered at the county level. Column 1 replicates the main findings from Table C.2 using a linear probability model. Column 2 estimates the effect of fracking on whether or not a case was filed in a given year using a logistic regression. Columns 3 and 4 show results for the number of cases filed using the Inverse Hyperbolic Sine (IHS) transformation and Poisson model, respectively. Similarly, in columns 5 and 6 the effect on number of charges filed is shown for both IHS and Poisson models.