Title
Essays on the Economics of Justice and Gender

Author
Sullivan, Briana Deann

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays on the Economics of Justice and Gender

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Economics

by

Briana Deann Sullivan

Committee in charge:

Professor Kate Antonovics, Co-Chair
Professor Gordon Dahl, Co-Chair
Professor Prashant Bharadwaj
Professor Craig McIntosh
Professor Sally Sadoff

2019

The Dissertation of Briana Deann Sullivan is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Co-Chair

University of California San Diego

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

VITA

2012	Bachelor of Arts, University of Florida
2015	Master of Arts, University of California San Diego
2012-2019	Teaching Assistant, University of California San Diego
2019	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

"Credit Line Use and Availability in the Financial Crisis: The Importance of Hedging," with Jose M. Berrospide and Ralf Meisenzahl, *Finance and Economics Discussion Series 2012-27,* Board of Governors of the Federal Reserve System (U.S.), 2012.

FIELDS OF STUDY

Major Field: Economics

ABSTRACT OF THE DISSERTATION

Essays on the Economics of Justice and Gender

by

Briana Deann Sullivan

Doctor of Philosophy in Economics

University of California San Diego, 2019

Professor Kate Antonovics, Co-Chair
Professor Gordon Dahl, Co-Chair

This dissertation consists of three papers on the economics of justice and gender. The first chapter pertains to the gender wage gap, and the second and third papers study the effects of criminal justice reforms.

My first chapter replicates and extends on Mulligan and Rubinstein (2008), which finds that the change in the female labor force composition significantly contributed to the narrowing of the gender wage gap. Their results, however, are obtained using the Current Population Survey (CPS). Therefore, I extend on their work by: (1) replicating their methodology using Census data, which is more representative of the US population, and (2) addressing changing male labor force

participation rates. Overall, my results continue to suggest that changing selection into the female labor force drives the convergence of the gender wage gap. But failing to account for changing male labor force participation rates overestimates convergence.

Chapter 2 studies how Alabama's shift from voluntary to presumptive sentencing guidelines affected sentence length. Using the National Corrections Reporting Program (NCRP), I find that after the guidelines became presumptive for a subset of property and drug offenses, average sentence length for affected offenders fell, with no statistically significant difference in effects across race. Using quantile regressions, I find that among affected property and drug offenders, the largest changes in total sentence length occur in the upper tail of the sentence length distribution.

Chapter 3 presents a descriptive analysis of the differential impacts of AB 109 (or Realignment) and Proposition 47 across California's counties. In October 2011, California implemented Realignment, which shifted the supervision responsibility for low-level offenders from the state to the county-level. Yet, I do not find a statistically significant relationship between the percent change in the prison and jail incarceration rates following AB 109. Then in November 2014, California implemented Proposition 47, which re-classified low-level felonies as misdemeanors and therefore reduced the jail population. Overall, I find that the fiscal impact of AB 109 may have been the largest for lower-income counties, with little evidence suggesting that Proposition 47 mitigated the fiscal impacts of AB 109 among these low-income counties.

Chapter 1

A Replication: Selection, Investment, and Women's Relative Wages Over Time

1.1 Introduction

Since the 1970s, there has been a striking convergence in the gender wage gap and a simultaneous increase in the female labor force participation rate. As seen in Figure 1.1, between 1970 and 2014, there was a 56 percent decrease in the gender wage gap among full-time full-year white, non-Hispanic men and women aged 26 to 55. During the same time period, while there were small fluctuations in the male full-time full-year labor force participation rate, women experienced an increase in participation from 28 percent to 51 percent. However, it is not evident that the convergence in the gender wage gap is due to a change in the relative wages of men and women, holding constant the composition of the labor force. Historically, women have had significantly lower labor force participation rates relative to men, so selection into the female labor force is likely. Additionally, because female labor force participation increased dramatically between 1970 and 2014, the changing composition of the female labor force may have contributed to the narrowing of the gender wage gap. If the selection rule for women has changed over time, the observed convergence in the gender wage gap may not reflect the true trend. Therefore, it is important to examine whether female selection into the labor force has changed over time and how women's relative wages have evolved holding constant the composition of the labor force.

This paper replicates Mulligan and Rubinstein (2008), henceforth referred to as MR, and extends on their paper by using Census data, rather than the Current Population Survey (CPS), and doing more to model male selection into the full-time full-year labor force. MR examine how changing female selection bias has contributed to the convergence of the gender wage gap. They employ two methods using repeated CPS cross sections to calculate a selection-corrected relative wage series for women between the 1970s and 1990s. The first method is the Heckman two-step estimator with marital status interacted with the number of children under the age of 6 as the exclusion restriction. Second, MR use identification at infinity to calculate the wage gap between men and women who are predicted to be strongly attached to the full-time full-year labor force, and therefore are subject to smaller selection bias. Overall, they find that women were negatively selected into the labor force in the 1970s and positively selected in the 1990s and conclude that much of the narrowing of the gender wage gap reflects this change in the composition of the female labor force. They attribute the convergence of the observed gender wage gap to the growing wage inequality within gender; an increase in the return to skill over time will attract more skilled women into the labor force and push less skilled women out.

A key component of the Heckman two-step method is the correct estimation of the first-stage labor force participation equation. Yet, because the CPS is a select sample that excludes individuals residing in group quarters, such as correctional institutions, and selection is nonrandom, the first-stage estimates and associated selection-corrected wage estimates are likely to be biased. Since the CPS is becoming more select over time due to the growing incarcerated population, using the CPS to estimate selection-corrected wages can also yield a misleading estimate of the change in the gender wage gap. In contrast, the Census is more representative of the U.S. population and includes individuals residing in group quarters.[1] However, due to limitations of the Census in recording the number of children for individuals living in group quarters, I am unable to use the full Census sample to address how the omission of individuals residing in group quarters from the CPS affects the Heckman two-step results. Thus, I replicate the Heckman two-step method using a modified Census sample.

MR further examine the effects of selection into the full-time full-year labor force by using individuals predicted to have strong labor force attachment to estimate a gender wage gap less influenced by selection bias. But the select nature of the CPS sample also presents a problem for identification at infinity because it requires the correct identification of demographic groups that have high labor force participation rates, and thus, for whom selection is likely to be a minor concern. Namely, when the data are missing for certain subpopulations, using the CPS may overestimate full-time full-year labor force participation for some demographic groups and, consequently, yield a more biased estimate of the gender wage gap relative to estimates obtained using the Census. In particular, because a larger percentage of men, relative to women, reside in group quarters, men's estimated wages are disproportionately affected by the omission of the group quarters population from the CPS. Furthermore, because the CPS is becoming more select over time, if the gender wage gap is estimated using a subset of men that appears to have strong labor force attachment, but the group is actually becoming more positively selected over time, then the CPS is likely to mis-estimate the level of the wage gap, particularly in later years, and therefore mis-estimate the convergence of the gender wage gap.

A limitation of MR, and this paper, is that by not modeling male selection in the Heckman two-step method, they do not address whether selection for men has changed. And while MR account for the selection of men into the full-time full-year labor force in identification at infinity,

[1] Group quarters are often defined as units with 10 or more individuals unrelated to the householder. Institutional group quarters refer to correctional institutions, mental institutions, and institutions for the elderly, handicapped, and poor. Non-institutional group quarters refers to military barracks, college dormitories, and rooming houses. Individuals living in group quarters are sampled as individuals; therefore, the number of children or other household related entries are not reported.

they do not address how changing selection rules among men can affect the convergence of the gender wage gap. In 1975, the wages for males with a high school degree were 80 percent of those earned by males with at least a college degree. In 2014, this number fell to 60 percent (Council of Economic Advisers 2016). This decline in wages for less-skilled men is likely to be influenced by selection into the labor force because of the fall in the employment of less-educated and low-wage males. Incarceration rates for men have also been rising rapidly; in 1990 the rate of incarceration was 564 per 1000,000 in the population and rose to 890 per 100,000 in 2014 (Council of Economic Advisers 2016). Consequently, because changing selection bias among men will bias the observed gender wage gap, this paper attempts to improve the identification at infinity selection equation for men by using the full Census sample, which includes individuals living in group quarters, and by accounting for the recent declines in male labor force participation rates. I also use a modified Census sample, meant to mimic the CPS, to replicate the MR identification at infinity results and assess whether differences between the results using the CPS and the full Census sample are driven by different sample universes or different datasets.

Section 2 of this paper discusses the relevant literature. Section 3 replicates MR using the same methodology and sample from the CPS. While there are some elements of MR's findings I am unable to replicate, like them, I find that after controlling for female selection into the full-time full-year labor force, the convergence of the gender wage gap disappears. Section 3 of the paper describes the use of Census data to replicate MR's methodology. When I control for female selection using the same exclusion restriction as MR for the Heckman two-step method, the gender wage gap again appears relatively constant over time. However, between the 1970s and 1990s, while the CPS results show a slight widening of the gender wage gap by 0.011 log points, the Census results display a small narrowing of the gap by 0.014 log points. I extend on the identification at infinity analysis by including the group quarters population in the sample and modifying how male selection into the labor force is modeled. I find slightly more convergence in the selection-corrected gender wage gap among individuals predicted to be strongly attached to the labor force when using Census data relative to CPS data. Nonetheless, MR's conclusion that changing selection into the female labor force contributes to the convergence of the gender wage gap does not appear to be sensitive to the dataset used. Section 4 concludes.

1.2 Literature Review

Researchers have provided a variety of reasons to explain the convergence of the gender wage gap, such as an increase in womens experience relative to that of men, the reversal of the gender education gap, and a shift into better paid occupations.[2] Because the proportion of women employed full-time full-year has changed over time in size and composition, changing selection bias is also likely to contribute to the narrowing of the gender wage gap. Yet, due to differences in the methodology and data used, papers studying the effects of selection bias have reached contradictory conclusions about the direction of the bias and the extent to which changing selection can explain the convergence of the wage gap. Namely, most papers suggest that selection into the female labor force was positive in the 1990s, but there is no consensus on whether selection was negative or positive in the 1970s.

Like MR, some papers use the CPS and Heckman two-step method, albeit with different exclusion restrictions, to estimate a selection-corrected gender wage gap. In contrast to MR, these papers control for both male and female selection into the labor force, yet it is difficult to find an exclusion restriction for men that affects labor force participation but not market wages. Blau and Beller (1988) use the 1972 and 1982 CPS to study earnings differentials by gender and race. Using nonlabor income, an indicator for whether an individual was at least 62 years of age, and the number of other family members between 18 and 64 years old as instruments, they find that selection into the labor force among white men was negative in 1971 but fell in magnitude in 1981 and that selection among white women became increasingly positive. Overall, they find that

[2]See Blau and Kahn (2017) for an extensive literature review on the gender wage gap.

between 1971 and 1981, the selection-corrected wages of white women rose relative to those of white men, in contrast to the constant trend in observed earnings differentials. Unlike Blau and Beller, but consistent with MRs findings, Jacobsen, Khamis, and Yuksel (2015) use 50 years of CPS data and the Heckman two-step method with marital status as an exclusion restriction, to find that selection into the female labor force was negative in the 1970s but became positive over time. Yet, similar to Blau and Beller, they find negative, but relatively constant, selection into the male labor force.

Another method used to study selection bias and the gender wage gap is the imputation of wages for those not in the full-time full-year labor force. To estimate a selection-corrected gender wage gap, wage imputations can non-parametrically control for the omission of non-employed women with high potential wages and non-employed men with low potential wages. While most papers find that selection into the female labor force was positive since the 1990s, Blau and Kahn (2006), using wage history and imputation methods, find that women were positively selected into the labor force in the 1980s and negatively selected in the 1990s, possibly due to changes in welfare laws that encouraged relatively low-skilled women to enter the labor force.

Despite the gender wage gap literature attempting to account for or impute the wages of the non-employed, it fails to consider the wages of individuals residing in group quarters, most of whom are men. Although Herrmann and Machado (2012) do not account for the wages of those in group quarters, they emphasize the importance of accounting for male selection into the labor force. The authors examine how cognitive ability affects selection into employment and find that differential female selection into the labor force over time cannot explain much of the narrowing of the gender wage gap. They show that young women became less positively selected between the 1960s and 1990s but that the relationship between ability and full-time full-year labor force participation disappears for older women. Men, on the other hand, appeared to positively select into employment. They conclude that selection bias due to ability is exaggerated for women and observed wages can be used to measure average female potential wages, while male selection may be important when measuring the gender wage gap.

Lastly, other authors have sought to expand on MR. They have found that the selection of women into the labor force in the late 1970s was positive, rather than negative, and that selection bias plays a smaller role, relative to MRs results, in the convergence of the gender wage gap. Bar, Kim, and Leukhina (2015) correct for inconsistent estimates of the wages of married women, resulting from omitted variable bias, by including spousal income as a determinant of female full-time full-year labor force participation in the first stage of the Heckman two-step method. They find that relative to MR's results, changing selection bias plays a less important role in the convergence of the gender wage gap. Rather, the decline in the market price gap paid on observables, such as experience and educational attainment, can explain a larger proportion of the narrowing of the gender wage gap. They also find that selection is positive in both the 1970s and 1990s. Schwiebert (2012) also builds on MR by addressing the possibility of inconsistent parameter estimates caused by the endogeneity of education, which can arise due to unobservable factors that affect wages, the probability of labor force participation, and educational attainment. The paper uses quarter of birth as an instrument for education and estimates womens wages over time using the Heckman two-step method. Because the instrument is not available in the CPS, the paper uses the 1980 Census and 2005 to 2010 American Community Survey (ACS) data but continues to limit the sample to individuals not residing in group quarters. The paper finds that selection has become more positive over time and was not negative in 1980.

Since it is difficult to find an appropriate exclusion restriction for males, this paper uses Census data to estimate a selection-corrected wage gap through identification at infinity, rather than the Heckman two-step method. I aim to contribute to the gender wage gap literature by using the Census and ACS to include individuals residing in group quarters in the first and second stages of the identification at infinity analysis. I also account for how declining male full-time full-year labor force participation rates can affect which groups of men have wages that are minimally affected by selection bias.

1.3 Replication

1.3.1 Data

Following MR, this paper uses the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS), also known as the March Supplement. As in MR's original paper, to increase sample size, this section creates two sets of repeated cross sections, each of which contain five years of survey data: 1976 to 1980 and 1996 to 2000 survey-years, or 1975 to 1979 and 1995 to 1999 work-years, respectively.

The sample consists of white, non-Hispanic individuals aged 26 to 55, not living in group quarters. MR classify individuals as full-time full-year if they worked at least 35 hours per week and at least 50 weeks per year last year. For log hourly wage estimation, they exclude the Armed Forces population, the self-employed, and individuals in agriculture or private household sectors. MR also exclude individuals with allocated earnings.

Hourly wages, denominated in year-2000 dollars, are calculated using annual earnings, usual hours worked per week, and weeks worked last year.[3] As a further sample restriction for log wage estimation, MR trim wages such that observations with an hourly wage under \$2.80 per hour are excluded. Next, they generate a full-time full-year log hourly wage distribution, which includes wages for white males of all categories of marital status and non-white, never-married men aged 18 to 65 (at the time income was received). Individuals below the 1^{st} percentile and above the 98^{th} percentile are excluded from the log wage regressions. Trimming drops most individuals whose wages were topcoded. While MR make their wage cutoffs publicly available, I am unable to generate the same values using the method described above. I obtain a higher minimum and maximum wage cutoff. My 1^{st} percentile cutoffs are often larger by at most \$1,

[3]Prior to 1976, the ASEC did not inquire about usual weekly hours worked and weeks worked in the last year; rather, hours and weeks worked are reported in intervals. Therefore, I use the data provided by Casey Mulligan which includes imputations for hours and weeks worked variables. See MR for a description of the imputation procedure.

and my 98^{th} percentile values are, on average, larger by approximately $9. I use my derived wage cutoffs to generate all results that use the wage-trimmed sample.

Following MR, I use six education categories: 0 to 8 years of schooling, 9 to 11 years of schooling, received high school degree, received college degree, and received advanced degree.[4] I calculate potential experience by taking the maximum of zero or age minus years of schooling completed minus 7. All calculations use person weights provided in the CPS.

Tables 1.1 and 1.2 contain summary statistics and are analogues of Table A.1 in MR. Table 1.1 compares the full CPS sample's descriptive statistics from MR to descriptive statistics from the replication. Means from MR and my replication are often identical, with differences likely due to rounding. Table 1.2 compares descriptive statistics for the full-time full-year CPS sample from MR and from the replication. While the differences, stemming from different wage cutoffs, are statistically significant, they are typically less than .002 in magnitude.[5]

1.3.2 Heckman Two-Step

1.3.2.1 Methodology

Following MR, this paper begins by using Heckman two-step estimation (see Gronau (1974) and Heckman (1979)) to adjust for the selection of women into the full-time full-year labor force. The first step of the Heckman procedure estimates a probit model, Equation (1.1), using the full sample of women with full-time full-year status as the dependent variable. To better model

[4]Prior to 1992, the CPS classified education by the years of schooling, rather than the highest degree received. Therefore, before 1992 high school graduates refers to individuals with 12 years of schooling, some college refers to 13 to 15 years of schooling, college refers to 16 to 17 years of schooling, and advanced degree refers to at least 18 years of schooling. Years of schooling are imputed for survey years 1992 to 2003 following Autor, Katz, and Kearney (2005).

[5]The largest difference between the MR and replicated summary statistics is in average Potential Experience-15 for 1975-1979. In Table 1.1, MR list the average potential experience for women as 5.073 years (Column 1), while the average potential experience under the replication is 5.067 years (Column 3). I cannot comment on statistical significance because MR do not list standard errors in the summary statistics. In comparing MR and the exact replication, these differences are unusual because all other averages have at most a difference in magnitude of 0.001. Nonetheless, they do not appear to significantly affect the marginal effects or coefficients on potential experience when exactly replicating MR.

selection, Heckman two-step estimation requires an exclusion restriction meant only to affect reservation wages not market wages. Following MR, the exclusion restriction is marital status interacted with the number of children under the age of 6.[6] The second step of the Heckman method, Equation (1.2), estimates a log wage least-squares regression using the full-time full-year sample of women and includes the inverse Mills' ratio estimated in the first stage.

$$P_t(\mathbf{Z}) \equiv Pr(L=1|\mathbf{Z}, g=1) = \Phi(\mathbf{Z}\delta_t) \tag{1.1}$$

$$w_{it} = \alpha_t + \mathbf{X}_{it}\beta_t + \theta_t\lambda(\mathbf{Z}_{it}\delta_t) + u_{it} \tag{1.2}$$

Let t denote time period, particularly the 1975-1979 and 1995-1999 work-years. Let $L=1$ denote full-time full-year labor force participation. Let g represent sex, such that 1 and 0 represent females and males, respectively. The row vector $\mathbf{X}$ contains demographic characteristics affecting market wages, namely educational attainment, marital status, region, and a potential experience quartic interacted with education indicators. Let $\mathbf{Z}$ denote a row vector that includes all elements of $\mathbf{X}$ and the exclusion restriction of marital status interacted with the number of children under the age of 6. $\Phi(\cdot)$ represents the cumulative distribution function of the standard normal distribution. Estimating Equation (1.1) yields $\hat{\delta}_t$, a vector of coefficients at time t. Overall, Equation (1.1) estimates the probability that a woman participates in the full-time full-year labor force at time period t, conditional on a set of demographic variables and the exclusion restriction.

Let w_{it} represent a female's log hourly wage at time t. Using estimates from the probit equation, an inverse Mills' ratio, $\lambda(\mathbf{Z}_{it}\hat{\delta}_t)$, is assigned to each observation. Let β_t be the vector of coefficients on demographic variables $\mathbf{X}$ and θ_t be the coefficient on the female inverse Mills'

[6] Heckman (1974) first used the number of children under the age of 6 as an instrument for women's reservation wages. The idea is that as the number of young children increases, a woman's reservation wage increases and the more likely she is to be out of the labor force. However, the number of children should have no effect on the mother's market wage. By interacting the number of children with marital status, MR take into account the fact that married women with children are more likely to stay out of the labor force than single women with children. Nonetheless, this may be an imperfect exclusion restriction because high earnings potential can affect a woman's decision to have children and marital status can be affected by wage structure.

ratio. The unobserved component of wages, the error term, is u_{it}. As in MR, there are some instances where marital status is excluded as a regressor in the log wage estimations because it is possible that an individual's marital status responds to wage structure.

1.3.2.2 Results

This section compares the estimates from my replication to the estimates from MR and re-summarizes MR's interpretation of the results. All tables referred to in this section contain two sets of columns: the first set lists results taken from MR, while the second set lists the replication results. Overall, in following MR's methodology and using their CPS data, I am able to achieve similar OLS and Heckman two-step estimates.

MR first compute OLS estimates of men's and women's log hourly wages using the full-time full-year sample of men and women. The OLS regression equation, run separately by sex, is similar to Equation (1.2), but the inverse Mills' term is omitted. Then, the Heckman two-step method is used to estimate selection-corrected log hourly wages for women. As described in Section 1.3.2.1, in the first stage of the Heckman two-step method, I use the full CPS sample to estimate the probit equation with full-time full-year female labor force participation as the dependent variable. Coefficients from probit regressions that include the exclusion restriction, marital status interacted with the number of children under the age of 6, are used to estimate an individual's inverse Mills' ratio, which is then included in the second-stage log wage regressions to control for selection into the labor force.

Tables 1.C.1, 1.C.2, and 1.C.3 are the equivalent of Table A.2 in MR, which lists the marginal effects from the probit and the OLS and Heckman two step-coefficients from the log hourly wage regressions. Table 1.C.1 compares the marginal effects derived from probit estimation with and without the exclusion restriction. There is at most a 0.001 (rounded) difference in magnitude between the marginal effects presented in MR and the replication. Tables 1.C.2 and 1.C.3 compare the coefficients from the regressions on log hourly wage, with and without the inverse Mills' ratio, for 1975-1979 and 1995-1999 work-years, respectively. Differences between the wage cutoffs I generate and those used in MR lead to small but statistically significant differences between the estimated log wage coefficients in the replication and in MR.

Next, MR estimate gender wage gaps using the OLS and Heckman two-step log hourly wage coefficients from Tables 1.C.2 and 1.C.3. I use what MR term as variable and fixed weights to estimate a male and female log hourly wage by substituting the weights into the estimated OLS and Heckman two-step regressions.[7] Weights are such that the average full-time full-year female and male are identical, except for sex.[8] Fixed weights have the average characteristics of full-time full-year women aggregated over 1975-1979 and 1995-1999, while variable weights differ by time period. In addition to unsuccessfully replicating the wage cutoffs used in MR, I am also unable to replicate the weights used to calculate the estimated gender wage gaps. Using the weights I generate, I then estimate an OLS gender wage gap for each time period by taking the difference between the OLS-estimated female and male log wages. Following MR, to estimate a Heckman two-step wage gap, I take the difference between the Heckman two-step-estimated female and the OLS-estimated male log hourly wages.

Table 1.3 mirrors Table I in MR, which illustrates how the gender wage gap has changed over time with and without controlling for selection into the full-time full-year labor force.[9] In Table 1.3, Panel A uses variable weights, while Panel B uses fixed weights. Columns titled "OLS" use the coefficients from the wage regressions without an inverse Mills' ratio to calculate the gender wage gap. Columns titled "Two-Step" use the coefficients from the wage regressions with an inverse Mills' ratio to calculate the gender wage gap. This yields the estimated gender wage gaps listed in Columns 1, 2, 4, and 5. The "Bias" columns (Columns 3 and 4) are calculated by taking the difference between the OLS-estimated wage gap and the Heckman two-step-estimated wage gap. These columns quantify selection bias since the OLS estimates do not use the inverse

[7]See Appendix 1.A for a description of how weights are generated and why variable and fixed weights are used.
[8]For weights, the inverse Mills' ratio is set to zero.
[9]Note that the estimated gender wage gaps in Table 1.3 are equivalent to the unexplained portion of the gender wage gaps from an Oaxaca decomposition.

Mills' ratio as a regressor to control for selection. Rows titled "Change" take the difference between the 1995-1999 estimated wage gap and the 1975-1979 estimated wage gap. They also include the difference of the 1995-1999 estimated bias and the 1975-1979 estimated bias.[10]

The differences between the wage cutoffs and weights used in the replication and in MR yield small but typically statistically significant differences in the OLS and Heckman two-step gender wage gaps listed in Table 1.3. For example, in MR the selection bias listed in Panel A increases by 0.162 log points (Column 3), while in the replication paper selection bias increases by 0.169 log points (Column 6). Because the differences between the MR and replication estimates are small, the interpretation of Table 1.3 remains unchanged. Overall, on average, the selection-corrected gender wage gap appears to have been relatively constant between the 1970s and 1990s, while the uncorrected (or OLS) gender wage gap has fallen over time. MR take this to suggest that part of the measured gender wage gap convergence stems from changes in the selection rules and composition of the female full-time full-year labor force. In particular, the "Change" rows of Column 4 in Table 1.3 show that between the 1970s and 1990s, the OLS estimates of the gender wage gap closed by 0.158 log points in Panel A and 0.143 log points in Panel B. These decreases in the OLS-estimated gender wage gaps support the convergence of the observed gender wage gap seen in Figure 1.1. Yet, controlling for selection bias, the point estimates of the gender wage gap widen over time, although the difference in the two-step estimates over time is insignificant. The estimates in the "Change" rows of Column 5 are now -.0011 and -0.025 in Panels A and B, respectively. Therefore, once the selection of females into the full-time full-year labor force is accounted for, the convergence of the estimated gender wage gap disappears. Column 6 shows that the estimated selection bias is negative for 1975-1979 (-0.078 in Panel A and -0.074 in Panel B), and for 1995-1999 the selection bias is positive (0.091 in Panel A and 0.094 in Panel B). Both

[10]The variance of the change in OLS estimates is calculated by summing the variance of the gender wage gap in the 1970s and the variance of the gender wage gap in the 1990s. The variance of the change in the two-step-estimated gender wage gap is calculated similarly. To calculate the variance of the bias, sum the variance in the OLS change and the variance in the two-step change. Note that this requires an assumption of independence. Following MR, standard errors are calculated using the non-parametric pairwise bootstrap method (1,000 replications) with a seed of 123.

differences are statistically different from zero. MR conclude that because the selection bias shifted from negative to positive between the 1970s and 1990s, the selection rule has changed over time.

MR summarize the partial “effects” of marital status and schooling by illustrating how the gender wage gaps would have evolved over time if all full-time full-year women fell into a particular marital or educational attainment demographic (if for example, all women were married or received college degrees). MR then compare how gender wage gaps for the average full-time full-year woman changed over time relative to the gender wage gaps of particular demographic subgroups. Table 1.4, which uses fixed weights, mirrors Table II in MR. As with Table 1.3, in Table 1.4, there small and sometimes statistically significant differences between the estimated gender wage gaps in MR and the replication, but the table’s interpretation remains the same. Columns 1 and 6 and 2 and 7 list the OLS-estimated gender wage gaps in 1975-1979 and 1995-1999, respectively. Columns 3 and 8 take the difference between the 1995-1999 and 1975-1979 OLS-estimated wage gaps. Comparing Panels A and B, Column 8 illustrates that conditional on being never-married, the OLS-estimated wage gap closed less relative to the OLS-estimated wage gap for the average woman (0.058 and 0.143, respectively). All other categories of marital status do not display such a small convergence in the OLS-estimated gender wage gap. Columns 4 and 9 list the difference between the 1995-1999 and 1975-1979 Heckman two-step-estimated wage gaps. They show that for all marital status categories other than never-married, the change over time in the selection-corrected estimate of the gender wage gap is not statistically different from zero. Columns 5 and 10, which estimate the change in selection bias over time, take the difference of the Heckman two-step change (Columns 4 and 9) and the OLS change (Columns 3 and 8). This is equivalent to calculating the amount of selection bias in each time period and then taking the difference between the bias calculated for 1995-1999 and the bias for 1975-1979. In Column 10, the estimated selection bias in the never-married row increased by 0.117 log points, relative to a maximum of 0.193 log points for the currently married group. A similar pattern can be seen in Panel C of Table 1.4. Namely, the change in the OLS-estimated gender wage gap (Column 8) and the change in selection bias (Column 10) are small for the advanced degree row relative to other educational attainment subgroups. Therefore, as discussed in MR, Table 1.4 illustrates that among demographic groups with high female labor force participation, such as the never-married or those with an advanced degree, because the estimated change in selection bias is relatively small, their wages are subject to less selection bias.

Despite the differences between MR and this paper in the wage cutoffs and weights used to generate Tables 1.3 and 1.4, the results continue to suggest that between the 1970s and 1990s, changing selection rules drove the convergence of the observed gender wage gap.[11,12] Namely, by controlling for female selection into the full-time full-year labor force, the selection-corrected gender wage gap appears to have been relatively constant during this time.

1.3.3 Identification at Infinity

MR use a simplified version of identification at infinity to further illustrate how, over time, there has been little change in the gender wage gap for demographic groups with strong full-time full-year labor force attachment. The Heckman two-step method requires an exclusion restriction, a variable that affects labor supply without affecting wages, to ensure that selection bias is not measured using the assumption of joint normality of errors (Heckman 1979). The purpose of identification at infinity, derived from Chamberlain (1986), Heckman (1990), and Schafgans and Zinde-Walsh (2002), is to estimate a wage equation using a sample selected to have strong labor force attachment, which equates to lower selection bias, without the use of an exclusion restriction. If the gender wage gap has converged due to changing selection bias, MR assume that measured wage changes should be smaller for demographic groups with high

[11] The wage cutoffs and weights used in MR are publicly provided by Casey Mulligan. When using the provided values, I am able to exactly replicate MR's results listed in Tables 1.3 and 1.4.

[12] In MR and my replication, I use pre-generated interactions for probit and log wage regressions. However, pre-generated interactions yield incorrect marginal effects and estimated gender wage gaps. Therefore, in Appendix 1.B, I replicate the Heckman two-step results from Section 1.3.2.2 using factor variables.

full-time full-year employment since selection bias is approximately zero in these groups. Note, there is a trade-off between sample size and the amount of selection bias: the higher the minimum labor force participation threshold, the fewer the number of demographic groups, and therefore individuals, there are that will meet this minimum. Additionally, identification at infinity does not estimate wage gaps for demographic groups that are excluded from log hourly wage estimation, so it is difficult to generalize the results from demographic groups with high labor force participation to groups with low labor force participation.

The first step of identification at infinity estimates a probit equation on full-time full-year status, separately by gender, using the vector of wage regressors, $\mathbf{X}_{it}$, from Equation (1.2) and omitting the exclusion restriction used in the Heckman two-step estimation. It yields $\widehat{P(\mathbf{X}_{it}\delta_g)}$, the estimated probability of participating in the labor force full-time full-year, where δ_g is a vector of coefficients on the probit regressors for sex g. The probit is run using only 1975-1979 data for both men and women. Following MR, I then use the probit estimates to select men and women who are full-time full-year and have characteristics such that the predicted probability of full-time full-year labor force participation exceeds some α near 1.

The next step is to calculate a selection-corrected gender wage gap using the demographic groups with high predicted labor force attachment:

$$(w_{it} - \mathbf{X}_{it}\hat{\beta}_t^{\infty}) = g_i\gamma_t + \varepsilon_{it}^{w}, \{i | P(\mathbf{X}_{it}\delta_g) \geq \alpha, L_{it} = 1\} \quad (1.3)$$

Let w_{it} be the log hourly wage, g_i is the sex of individual i (1 being female and 0 being male), and γ_t is the coefficient on sex at time t. Let $L_{it} = 1$ indicate that individual i at time t is participating in the labor force full-time full-year. $\mathbf{X}_{it}\hat{\beta}_t^{\infty}$ is the predicted wage using the vector of demographics $\mathbf{X}$ described in Section 1.3.2.1 (Heckman 1990; Schafgans and Andrews 1998). The coefficients in the vector β_t^{∞} are estimated using OLS and are the same for males and females; sex is not interacted with all other regressors but is included as a regressor. The predicted wage is derived from the complete full-time full-year sample used in the Section 1.3.2. Note that using the OLS coefficients ignores whether the coefficients themselves are affected by selection bias. Denote the error term as ε_{it}^{w}.

Following MR, I estimate γ_t in two steps. First, I take the difference between the measured log hourly wage, w_{it}, and the predicted log hourly wage, $\mathbf{X}_{it}\hat{\beta}_t^{\infty}$. Then I regress this difference, the left-hand side of Equation (1.3), on sex. The coefficient on sex, $\hat{\gamma}_t$, is considered the estimated gender wage gap.

Figure 1.2, which is equivalent to Figure V in MR, plots the estimated gender wage gap, $\hat{\gamma}_t$, using different threshold values for α. Because I am unable to replicate the wage cutoffs used in MR, as in Section 1.3.2, I use the wage cutoffs I generated following MR's wage trimming method to create the wage sample for the log wage regressions. As a result, there are small but statistically significant differences between Figure 1.2 and Figure V.[13] Regardless of the threshold, each $\hat{\gamma}_t$ in the replication is larger relative to $\hat{\gamma}_t$ in MR (see Figure 1.C.1).[14] Additionally, for each threshold, relative to the convergence illustrated in MR, there is less convergence in the replication's estimated wage gaps between 1970 and 2000. For example, in the replication the estimated gender wage gap for the 80 percent threshold converges by 0.004 log points, while the estimated gender wage gap in MR narrows by 0.02 log points.

Nonetheless, as in MR, Figure 1.2 illustrates that among demographic groups with high predicted full-time full-year labor force participation, the estimated gender wage gap trends little over time. This is especially true for groups with predicted labor force participation of 70 percent and 80 percent. Recall that as the threshold for the predicted probability of labor force participation falls, the amount of selection bias increases. This can explain why in Figure 1.2, there is more of a trend in the gender wage gap among individuals with at least a 60 percent predicted probability of working full-time full-year.

[13] When using the wage cutoffs provided by MR, there appears to be no significant differences between Figure 1.2 and Figure V in MR.

[14] The 80 percent threshold appears to be most sensitive to the change in wage cutoffs.

To address the concern that endogeneity may bias the probit and wage regression results, since marital status can be affected by wage structure and may have changed over time, MR include an x-marked series in Figure 1.2 that uses all regressors other than marital status to estimate the gender wage gap. However, a 50 percent threshold is used to ensure a large enough sample size of females. Therefore, MR expect some of the trend in the x-marked series to be a result of changing selection bias. Nevertheless, because the trend in the x-marked series is small, MR conclude that the convergence of the gender wage gap does not depend on the use of marital status as a regressor.

Overall, as in MR, the unmarked line denoting the uncorrected gender wage gap shows a convergence of over 0.17 log points. The square-marked series, denoting the estimated gender wage gap between men and women predicted to have a probability of working full-time full-year of at least 80 percent, shows that the gender wage gap has not changed substantially over time. As a result, MR conclude that the convergence in the observed gender wage gap is due to changing selection into the female full-time full-year labor force.

1.4 Extensions using Census Data

This paper extends on MR in two ways. First, to verify whether MR's results are sensitive to the dataset used, I will replicate MR's Heckman two-step and identification at infinity analyses using a sample of Census data that mimics the CPS. Second, to better model selection into the labor force by using a more complete sample, I will replicate the identification at infinity analysis using the entire Census sample.[15] While MR use the 1975-1979 CPS samples estimate a probit on full-time full-year labor force participation to select both men and women with high predicted labor force participation, in order to account for declining male labor force participation, I use the 2010 ACS to estimate the first-stage coefficients for males. Following MR, I use the 1980 Census

[15] This paper uses the 2010 American Community Survey (ACS), the 2000 and 1990 5% Census samples, the 1980 5% Census sample, and the 1970 1% state fm1 Census sample, available from IPUMS-USA (Ruggles et al. 2015).

to estimate the probability of labor force participation for women.

Unlike the Census, the CPS excludes individuals residing in group quarters, such as the incarcerated.[16] The select nature of the CPS sample makes it difficult to estimate selection-corrected wages for both men and women. The issue is particularly relevant for men, who are likely to be disproportionately affected by the CPS' omission of individuals residing in group quarters. The select nature of the CPS is also problematic for correctly estimating the change in the gender wage gap over time. For the identification at infinity analysis, if the change over time in the gender wage gap is being estimated using a subset of men that appears to have strong labor force attachment, as suggested by the CPS, but the group is actually becoming more positively selected over time, then all else equal, the second-stage wage estimates will mis-estimate both the level and convergence of the gender wage gap. On the other hand, the Census is more representative of the U.S. population and can better model the selection of men and women.

1.4.1 Heckman Two-Step

1.4.1.1 Data

This section compares the OLS and Heckman two-step results from the 1980 and 2000 Censuses to the 1975-1979 and 1995-1999 work-year CPS samples, respectively. I also extend the Heckman two-step analysis to later years using the 2010 ACS. MR hypothesize that wage inequality within gender drives wage equality across genders. Since 2000, wage inequality within gender (the difference between the 90^{th} and 10^{th} percentiles of log hourly wages) has steadily increased, and the observed gender wage gap has closed further (see Figure 1.C.2). If selection bias has continued to contribute to the narrowing of the gender wage gap, convergence in the selection-corrected wage gap should remain small (or negligible).[17]

[16] As of 1983, among those residing in group quarters, only individuals residing in non-institutional group quarters were sampled in the CPS.

[17] As a comparison to the results from the 2010 ACS, I also include the Heckman two-step results estimated using the 2005-2009 and 2010-2014 (work-year) CPS samples in Appendix 1.C. Summary statistics for these CPS samples

Ideally, by using the Census I would be able to estimate the Heckman two-step selection equation using a more complete sample of women. However, I am unable to use the Heckman two-step model with the full Census sample because individuals living in group quarters are listed as having no children since they are sampled as individuals and not households. All Heckman results are therefore estimated using a "CPS-Similar" sample, which is meant to match the CPS sample used by MR. Although imperfect, the purpose of the "CPS-Similar" sample is to verify the CPS Heckman results using another dataset.

As described in Section 1.3.1, the sample includes white, non-Hispanic individuals aged 26 to 55, not residing in group quarters. I attempt to match education categories across the CPS and Census.[18] Following MR, I generate wage cutoffs using the Census' wage distribution for full-time full-year men (excluding those in group quarters) and estimate variable and fixed weights using the average full-time full-year woman. All calculations use person weights provided in the Census files.

Because the CPS is meant to be a labor market survey, the Bureau of Labor Statistics excludes members of the Armed Forces, namely active duty members, when reporting the U.S. unemployment rate. Weeks worked last year and usual hours worked per week (last year) for individuals in the Armed Forces are not recorded until 1990 (for both males and females), because in 1990, the CPS changed its definition of the sample universe from civilians to persons. Except for 1968, no women are listed as Armed Forces' members until 1988. Therefore, the "CPS-Similar" sample includes male members of the military for the 1980 and 2000 Censuses and the 2010 ACS; it includes female members of the military for the 2000 and 2010 samples. For selection into the labor force, all individuals in the military are considered as not working full-time full-year in the 1980 Census. Their full-time full-year status in the 2000 Census and 2010 ACS

are listed in Table 1.C.15.

[18]The Census uses highest grade completed. Detail varies by year of the survey, but I apply the same restrictions from MR to generate my education indicators. To calculate potential experience in the 1990 and 2000 Censuses, I first impute years of education for each individual using data from Autor, Katz, and Kearney (2005) and then calculate experience using the definition described in Section 1.3.1.

is determined by the weeks per year worked and hours per week worked reported. As in MR, individuals in the military are dropped for the wage regressions.

Summary statistics using the "CPS-Similar" sample are listed in the first four columns of Tables 1.5, 1.6, and 1.7 (which correspond to the 1980 Census, the 2000 Census, and the 2010 ACS, respectively). Columns 1 and 2 of each table list the descriptive statistics for the full "CPS-Similar" sample, and Columns 3 and 4 list the descriptive statistics for the full-time full-year sample. A comparison of the summary statistics in Table 1.1 to those in Tables 1.5 and 1.6 shows that, on average, a greater fraction of men and women are employed full-time full-year in the CPS than in the Census.[19] Additionally, among all individuals in the "CPS-Similar" sample, average potential experience increased between 1980 and 2000, while in the CPS, average potential experience decreased. For example, in Columns 3 and 4 of Table 1.1, average potential experience for women in the CPS fell by 0.54 years (4.527 - 5.067). In Column 1 of Tables 1.5 and 1.6, average potential experience for women in the "CPS-Similar" sample increased by 0.394 years (4.919 - 4.525). For the full-time full-year sample in the CPS, potential experience increases over time. Between the 1970s and 1990s, Columns 5 and 6 of Table 1.2 show that the average potential experience for a full-time full-year woman in the CPS increased by 0.034 years (4.136 - 4.102). There are even larger increases in average potential experience in the full-time full-year "CPS-Similar" sample. Column 3 of Tables 1.5 and 1.6 show that full-time full-year women increased their average potential experience by 0.531 years (4.436 - 3.905).[20] Average education is similar across the CPS and the "CPS-Similar" sample; although there are small but statistically significant differences between the educational attainment distributions.

[19]According to the BLS, differences between the CPS and Census in employment rates stem from differences in the interview environment (or lack thereof), specificity of the survey questions, and the definition of reference week. Additionally, as will be addressed in Section 1.4.2.1, there appear to be large differences between the CPS and Census in the full-time full-year labor force participation of women with at least a college degree, primarily due to how teachers in the Census report their weeks worked per year.

[20]The large differences between average potential experience in the CPS and Census may stem from the fact that the 1976 to 1980 survey-year CPS data is aggregated. When using the 1980 and 2000 CPS samples, the changes in average potential experience are of the same direction relative to those in the Census. Potential experience is also sensitive to how years of education is calculated. There do not appear to be any large differences in the CPS and Census age distributions.

1.4.1.2 Results

Following MR, I begin by estimating OLS and Heckman two-step log hourly wages. In the first stage of the Heckman two-step method, I model the selection of women into the full-time full-year labor force using the exclusion restriction of marital status interacted with the number of children under the age of 6 (see Section 1.3.2.1). The estimated first-stage coefficients are then used to estimate women's inverse Mills' ratio, which is included in the second stage log wage regressions. Table 1.C.4 lists the marginal effects from the probit estimation, and Table 1.C.5 lists the OLS and Heckman log hourly wage coefficients. Comparing the marginal effects listed in Tables 1.C.1 and 1.C.4, the only striking difference is that, relative to women with some college, women with at least a college degree in the CPS are more likely to work full-time full-year, while in the 1980 and 2000 Censuses they are less likely to work full-time full-year. Nonetheless, the difference in the sign of the marginal effects across samples does not seem to qualitatively impact the log wage regression coefficients, namely having at least a college degree is positively associated with log hourly wages.

I then use the coefficients from the OLS and Heckman two-step log wage regressions and the variable and fixed weights discussed in Section 1.3.2.2 to estimate the gender wage gaps listed in Tables 1.8 and 1.9.[21] Table 1.8 compares the estimated gender wage gaps between 1980 and 2000, and Table 1.9 compares the estimated gender wage gaps between 1980 and 2010. Panel A uses variable weights, while Panel B uses fixed weights. In both Panels A and B of Table 1.8, the "CPS-Similar" sample shows that the OLS-estimated gender wage gap converges more between 1980 and 2000 when using the Census rather than the CPS (see Table 1.3), a statistically significant difference. Panel A of Table 1.8 also shows a small, but statistically significant, narrowing of the selection-corrected gender wage gap by 0.014 log points (Column 2). On the other hand, Panel A of Table 1.3, which uses the CPS, shows a statistically insignificant

[21] Following MR, standard errors are calculated using the non-parametric pairwise bootstrap method (1,000 replications) with seeds of 123, 456, 78, and 91.

widening of the selection-corrected gender wage gap by 0.011 log points (Column 5). Table 1.9, which compares the estimated gender wage gaps across 1980 and 2010, tells a similar story.[22]

As illustrated by MR using the CPS, the results using Census data continue to suggest that much of the convergence in the gender wage gap can be attributed to changing selection bias. Both the CPS and Census results show a change in the selection rule over time, negative in the late 1970s and positive in the late 1990s. Comparing the "Bias" column in Tables 1.3 and 1.8 (Columns 6 and 3, respectively) the changes in the selection bias between 1980 and 2000 are of the same sign and similar magnitude, namely over 0.16 log points in both Panels A and B. Note, estimated female selection into the full-time full-year labor force remains positive when using the 2010 ACS (see Column 3 of Table 1.9).

Comparing the "partial effects" of marital status and educational attainment on the estimated gender wage gap, Table 1.10, which uses 1980 and 2000 Census data, yields similar qualitative results as in Table 1.4, which uses CPS data. Within their respective panels, the never-married and advanced degree groups in Table 1.10 have some of the smallest values for the change in the OLS-estimated wage gap (Column 3). In Column 5, these two groups also have smaller values for the change in selection bias relative to the other marital status and educational categories. For example, the change in selection bias for the never-married group is 0.125 log points relative to a maximum of 0.185 log points for the currently married group or relative to 0.164 log points for the average full-time full-year woman. Similar to MR, this suggests that for particular demographic groups that have a smaller estimated bias relative to that of the average full-time full-year woman, the estimated gender wage gap is less affected by selection bias.[23]

Overall, using the "CPS-Similar" sample and the Heckman two-step method continues to illustrate that once female selection into the full-time full-year labor force is accounted for,

[22] Table 1.C.17, in Appendix 1.C, uses the 2006 to 2010 and 2011 to 2015 CPS samples to draw a comparison between the Census and CPS in later years. Marginal effects from the probit estimation and log hourly wage coefficients are listed in Table 1.C.16.

[23] Using the Census and ACS, Table 1.C.6 compares the "partial effects" of marital status and educational attainment between 1980 and 2010. These can be compared to the "partial effects" estimated using the 2006 to 2010 and 2011 to 2015 CPS samples in Table 1.C.18.

the convergence in the observed gender wage gap disappears. Namely, both the CPS and the Census' "CPS-Similar" samples suggest that much of the narrowing of the gender wage gap is due to the changing selection of women into the labor force. The primary difference between the results obtained using the Census and the CPS, which will also be seen in the identification at infinity results below, is that a small and sometimes statistically significant narrowing of the selection-corrected gender wage gap occurs when using the Census sample; whereas a small widening in the wage gap is obtained when using the CPS.[24]

1.4.2 Identification at Infinity

Recall that by estimating a wage regression using a sample of individuals predicted to have strong full-time full-year labor force attachment, MR use identification at infinity to illustrate how the estimated gender wage gap trends over time for demographic groups that experience minimal selection bias. The first step estimates a probit on full-time full-year employment status. Following MR, I use the estimated probit to select a sample of individuals with high predicted labor force participation, and I then use the sample to estimate a selection-corrected gender wage gap. I first use the "CPS-Similar" sample, discussed in Section 1.4.1.1, to replicate the CPS identification at infinity results. Then I extend the analysis to the full Census sample. The first step determines whether any differences between the results using the full Census sample and CPS sample are driven by using different datasets rather than different sample universes.

Unlike the Census, the CPS excludes individuals residing in group quarters, so estimating the selection equation using the CPS may overestimate full-time full-year labor force participation for certain demographic groups. While this is true for men and women, men's estimated wages

[24] This difference might be a result of the mismeasurement of usual hours worked in the Census, since unlike the CPS, the Census does not use an interviewer to assist with questions. Baum-Snow and Neal (2009) find that in the 1980s there was a statistically insignificant difference between hourly wages from the CPS and Census. However, for white men and women, the 2000 Census yields larger hourly wages relative to hourly wages from the 2000 CPS. This difference is statistically significant, and women appear to have a larger discrepancy between hourly wages calculated using the Census relative to the CPS.

are more likely to be affected due to the CPS's selection rules. In particular, a larger percentage of men, relative to women, reside in group quarters. Additionally, when using the CPS, if the first stage of identification at infinity yields a subset of men that appears to have strong labor force attachment, but this group is actually becoming more positively selected over time, then, all else equal, using the CPS will mis-estimate the convergence of the gender wage gap.

Figure 1.C.3 depicts male and female full-time full-year labor force participation rates by educational attainment. It illustrates the differences in labor force attachment across the CPS and full Census samples and demonstrates the increasingly select nature of the CPS. Comparing the 1975-1979 CPS to the 1980 full Census sample and the 1995-1999 CPS to the 2000 full Census sample, Figure 1.C.3 shows that for each education category, there is a larger fraction of men employed full-time full-year in the CPS relative to the Census. The difference in labor force participation rates is particularly pronounced among less-educated men and the differences between the CPS and Census samples increase between 1980 and 2000. For example, among men with less than 8 years of schooling, in 1980, the fraction of men employed full-time full-year is 8 percent larger in the CPS relative to the Census, and in 2000, the fraction of men employed full-time full-year is 21 percent larger in the CPS relative to the Census. I, therefore, estimate the selection equation in the first stage and the wage regression in the second stage over a sample that includes individuals residing in group quarters.

To control for changing selection rules, MR use the 1975-1979 CPS data to generate the selection equation for both men and women. This estimated selection equation is used for all time periods; namely, the probit coefficients are the same for each set of years. Therefore, demographic groups selected for the wage regression in Equation (1.3) are predicted to have strong labor force attachment in the 1970s and are assumed to continue to have strong attachment. For women, using 1975-1979 CPS data makes sense because groups that have high labor force attachment in later years may not have been strongly attached in the 1970s. Using later years to estimate the selection equation would introduce selection bias into earlier estimates of the gender wage gap.

However, the full Census sample shows that male full-time full-year labor force participation has decreased from 73 percent to 68 percent between 1980 and 2010 (see Tables 1.5 and 1.7). Therefore, if men are increasingly and non-randomly exiting the labor force, in the 2010 ACS, some subsets of men who are predicted to have strong labor force attachment when using the 1980 Census will have estimated wages subject to more selection bias relative to subsets of men with consistently strong labor force attachment over time. As a result, the selection-corrected wage gap in the 2000s, and therefore the trend in the gender wage gap over time, are more likely to be mis-estimated.

As in Section 1.3.3, the first stage of identification at infinity estimates a probit on full-time full-year labor force status and the coefficients are used to select demographic groups predicted to have strong labor force attachment. Namely, subgroups of men for whom $P(\widehat{\mathbf{X}_{it}\delta_{g=0}}) \geq \alpha$ (where $\mathbf{X}_{it}$ is the vector of wage regressors, δ_g is a vector of coefficients on the probit regressors for sex g, and α is some threshold near 1) are used to estimate the gender wage gap in the second stage. MR use 1975-1979 CPS data (or equivalently the 1980 Census) to estimate the probit coefficients. Suppose there are men in particular demographic groups whose predicted labor force participation, estimated using 1980 data, is greater than α between 1980 and 2010. If, however, men in these subgroups are non-randomly exiting the labor force over time and I re-estimate their labor force participation using the 2010 data, then there may be too few men in this subgroup classified as full-time full-year, such that $P(\mathbf{X}_{i,\widehat{2010}\delta_{g=0}}) < \alpha$. To see how this affects the estimated gender wage gap, consider the following wage equations for men in 1980 and 2010:

$$w_{i,1980} = \mathbf{X}_{i,1980}\beta_1 + u_{i,1980} \tag{1.4a}$$

$$w_{i,2010} = \mathbf{X}_{i,2010}\beta_2 + u_{i,2010} \tag{1.4b}$$

Log hourly wages in 1980 and 2010 are denoted by $w_{i,1980}$ and $w_{i,2010}$, respectively. Let β_1 and β_2 represent the OLS coefficients on the vector of wage regressors $\mathbf{X}$ and $u_{i,1980}$ and $u_{i,2010}$ represent the error terms. Equations (1.5a) and (1.5b), below, are the conditional expectations of log hourly wages in 1980 and 2010, respectively.

$$E(w_{i,1980}|\mathbf{X}_{i,1980}, select_{1980}) = \mathbf{X}_{i,1980}\beta_1 + E(u_{i,1980}|\mathbf{X}_{i,1980}, select_{1980}) \tag{1.5a}$$

$$E(w_{i,2010}|\mathbf{X}_{i,2010}, select_{2010}) = \mathbf{X}_{i,2010}\beta_2 + E(u_{i,2010}|\mathbf{X}_{i,2010}, select_{2010}) \tag{1.5b}$$

The sample selection rules are defined as:

$$select_{1980} = (L_{i,1980} = 1, P(\mathbf{X}_{i,1980}\delta_{g=0}) \geq \alpha)$$

$$select_{2010} = (L_{i,2010} = 1, P(\mathbf{X}_{i,2010}\delta_{g=0}) \geq \alpha)$$

The selection rules indicate that male i at time t (1980 or 2010) is participating in the labor force full-time full-year and is predicted to have labor force participation greater than or equal to α. Assume that $\delta_{g=0}$ is estimated using 1980 data. If men are becoming more positively selected over time, then

$$E(u_{i,1980}|\mathbf{X}_{i,1980}, select_{1980}) < E(u_{i,2010}|\mathbf{X}_{i,2010}, select_{2010})$$

Namely, in the second stage of identification at infinity, men's estimated log hourly wages in 2010, relative to 1980, are subject to more selection bias. Consequently, both the gender wage gap and trend in the wage gap will be mis-estimated. If, instead, I estimate the first stage using 2010 data, the probit coefficients will select a group of men whose wages are less affected by selection bias over the entire sample period. Therefore, to address changing selection rules for men and women, I will use the 2010 ACS to estimate the selection equation for males, and I will continue to use the 1980 Census to model selection for females.

1.4.2.1 Data

The identification at infinity results illustrated in Figure V of MR and Figure 1.2 of this paper use CPS data from 1971 to 2000. To extend my analysis to the Census, I use the 1970 through 2000 Censuses. I add in the 2010 ACS to emphasize the fact that male full-time full-year labor force participation is declining over time. I estimate selection-corrected wage gaps using both the "CPS-Similar" sample and the full Census sample.[25] The full Census sample includes individuals residing in group quarters and those in the Armed Forces. Following MR, for the full Census sample, I generate wage cutoffs using the Census' wage distribution for full-time full-year men. For the "CPS-Similar" sample, I use the wage cutoffs described in Section 1.4.1.1.

MR use CPS data aggregated over 5-year intervals to estimate the identification at infinity gender wage gaps. I, instead, use 10-year intervals to compare the CPS and Census results. For example, I compare the estimated gender wage gaps from the 1980 Census and the pooled 1975 to 1984 (work-year) CPS samples. Because I use the 2010 ACS, I also include CPS data from 2004 to 2015 to the 1964 to 2003 survey-year data provided by Mulligan and Rubinstein.[26] In MR and Section 1.3.3, the 1975-1979 CPS data is used to estimate the first-stage probit coefficients for selecting individuals with high predicted labor force attachment. To compare the CPS and Census selection-corrected gender wage gaps, I instead, re-generate the CPS results using 10-year intervals; namely, I use the 1975-1984 CPS samples to estimate the probit coefficients.

Summary statistics using the 1980 and 2000 full Census samples and the 2010 full ACS sample are listed in the last four columns of Tables 1.5, 1.6 and 1.7, respectively. The "CPS-Similar" and full Census samples are different in terms of the sample over which I estimate the probability of participating in the full-time full-year labor force and in terms of the sample used for the log wage regressions. Comparing the first and last four columns within each table,

[25]For both samples, I impute hours per week and weeks per year worked in the 1970 Census and weeks per year worked in the 2010 ACS following Autor, Katz, and Kearney (2005).

[26]The 2004 to 2014 ASEC data is provided by IPUMS-CPS (Flood et al. 2015). Following Autor, Katz, and Kearney (2005), to calculate potential experience, I impute the highest grade attained.

including individuals residing in group quarters to create the full Census sample only yields small changes to the descriptive statistics relative to those of the “CPS-Similar” sample. The last two rows of Tables 1.5, 1.6, and 1.7 list the percentage of the full Census sample that reside in group quarters. Comparing Table 1.5 to Tables 1.6 and 1.7, Column 6 indicates that since 1980 the percentage of men in group quarters has increased from 1.2 percent to 1.9 percent, with much of the increase driven by incarceration.

1.4.2.2 Results

I begin this section by describing adjustments I make to the definition of the full-time full-year labor force. Then, I show how the selection-corrected gender wage gaps estimated using the full Census compare to those estimated using the CPS. When I account for declining male labor force participation rates and use the full Census sample, rather than the CPS, to estimate a gender wage gap among demographic groups with strong predicted full-time full-year labor force participation, I find relatively more convergence in the wage gap. Yet, as in MR, the narrowing of the selection-corrected gender wage gaps is much smaller relative to the uncorrected gender wage gap, which suggests that changing selection among women contributes to the observed convergence.

The highest full-time full-year labor force participation threshold that MR use is 80 percent, meaning that MR estimate a wage regression on individuals who have a predicted probability of participating in the labor force of at least 80 percent. When using the “CPS-Similar” sample and the full Census sample, I am unable to construct a graph of the gender wage gap over time, like Figure 1.2, for groups with strong predicted labor force attachment. Often, the estimated gender wage gap at each time point is zero, not because the wage gap is zero, but because there are no women who satisfy the criterion of having an estimated probability of full-time full-year labor force participation above 80 percent.

One possible explanation for lacking women with a high predicted probability of full-time full-year labor force participation is that women have lower participation rates in the Census relative to the CPS. To examine why I am unable to estimate a wage gap for large values of α, I compare women's full-time full-year labor force participation rates by educational attainment across the CPS and Census. I find that in the CPS, relative to the "CPS-Similar" sample, a greater fraction of women with at least a college degree are classified as full-time full-year; yet the two samples share a similar fraction of women with less than a college degree working full-time full-year. Since there are few highly-educated women classified as full-time full-year in the Census, having a college degree negatively predicts labor force participation (see Table 1.C.4). Despite other demographic variables positively predicting female labor force participation, such as potential experience, the marginal effect of educational attainment is either negative or small and positive; therefore, no female in the Census has an estimated probability above 80 percent.

The differences between the CPS and Census in labor force participation rates appear to be caused by discrepancies in the distributions of weeks worked for women who work as teachers or professors.[27] While not as apparent, male teachers and professors in the Census also have a different distribution of weeks worked per year relative to their counterparts in the CPS. Figures 1.C.4 and 1.C.5 illustrate how the distribution of weeks worked per year differ between the CPS and "CPS-Similar" samples for women and men who are classified as teachers or professors and for all other occupations. The figures indicate that individuals in non-teaching occupations in the Census and CPS share a similar distribution of weeks worked per year. However, Figure 1.C.4 shows that 20 percent of female teachers in the 1980 Census, compared to 53 percent in the 1976-1980 CPS samples, worked at least 50 weeks per year. To address the discrepancy in weeks worked per year for the identification at infinity analysis, I classify teachers in both the CPS and Census as full-time full-year if they work at least 35 hours per week and at least 36

[27] The difference in the distribution of weeks worked between the CPS and Census may be caused by how teachers record their weeks worked in the Census relative to the CPS. The CPS first asks the individual to consider all the jobs they worked during the past year and any weeks for which they worked even a few hours. Then the survey asks how many weeks they worked over the past year (including paid/sick leave). On the other hand, while the nature of the weeks worked question in the Census is similar, there are no prior questions that encourage respondents to think in detail about their work history.

weeks per year (an academic year is often 180 days or 36 weeks excluding winter, spring, and summer breaks).

To illustrate how modifying the definition of the full-time full-year labor force increases labor force participation rates among highly educated women, Figures 1.C.6 and 1.C.7 graph the percentage of women classified as full-time full-year by educational attainment. The graphs on the left correspond to the original full-time full-year definition of working at least 35 hours per week and at least 50 weeks per year. The graphs on the right correspond to the modified definition of labor force participation for teachers or professors and the original definition for all other occupations. The graphs show that when teachers or professors working at least 36 weeks per year are classified as full-time full-year, women's labor force status in the "CPS-Similar" sample becomes more comparable to the CPS. Namely, in Figure 1.C.6, the difference in full-time full-year labor force participation for women with at least a college degree between the 1975-1979 CPS and 1980 "CPS-Similar" sample went from 10 percentage points to 1 percentage point. In Figure 1.C.7, which compares the 1995-1999 CPS samples and the 2000 "CPS-Similar" sample, the difference falls from 7 percentage points to 3 percentage points.

After classifying teachers who work at least 36 weeks per year as full-time full-year, I am able to graph the gender wage gap time series, for all values of α, following the methodology in Section 1.3.3. I estimate the first stage of identification at infinity, a probit on full-time full-year labor force participation, to select demographic groups with high predicted labor force participation. The marginal effects on the probability of labor force participation, calculated using the "CPS-Similar" and full Census samples, are listed in Table 1.C.7. Next, I estimate a selection-corrected gender wage gap by regressing the difference of observed and predicted wages on sex using the sample of demographic groups with strong predicted labor force attachment

Figure 1.3 plots γ_t, the coefficient on sex estimated by Equation (1.3), using the full Census sample. The 1980 Census and the 2010 ACS are used to model female and male selection into the full-time full-year labor force, respectively. Figure 1.3, like Figure 1.2, illustrates that the convergence of the gender wage gap among those with high predicted labor force attachment is smaller relative to the narrowing of the uncorrected wage gap. Yet, as I discuss below, the selection-corrected wage gaps estimated using the full Census sample converge more relative to those estimated using the CPS. As in MR, convergence decreases as the threshold increases, suggesting that wages are subject to less selection bias as α increases.

To study the differences in convergence across samples, I separate the analysis into three components. First, I compare the identification at infinity results from the CPS and the "CPS-Similar sample, using 1980 (or 1975-1984) data to estimate both the male and female first-stage selection equations. Since the "CPS-Similar" sample is meant to mimic the CPS, the first step illustrates the effects of moving to a different dataset while using a sample with characteristics similar to those of the CPS. Second, to analyze how including individuals residing in group quarters affects the convergence of the wage gap, I compare the results from the "CPS-Similar sample to the results from the full Census sample, using the 1980 Census to model male and female selection into the labor force. Third, I examine how modifying male selection affects the estimated gender wage gaps. I compare results from the full Census sample using the 1980 Census to model female and male selection to results from the full Census sample using the 1980 Census for female selection and the 2010 ACS for male selection.

Differences between the CPS and "CPS-Similar" samples may stem from the type of demographic groups predicted to have strong full-time full-year labor force participation, particularly among women (see Tables 1.C.8, 1.C.9, 1.C.10, and 1.C.11 in Appendix 1.C). For instance, using the CPS, women predicted to have full-time full-year labor force participation of at least 70 percent have at least a high school degree. In contrast, using the Census samples, only women with at least a college degree are predicted to have strong labor force attachment in each year (even after modifying the definition of full-time full-year).[28]

[28]Note that the demographic groups for identification at infinity are not solely defined by one dimension, such as educational attainment or marital status. Rather they are defined by marital status x education x potential experience x region. Therefore, not all women with at least a college degree, for example, are predicted to have strong labor force attachment because having a college degree does guarantee that $P(\mathbf{X}_{it}\hat{\delta}_g)$ is greater than α.

To further illustrate differences in results across datasets, Figure 1.4 graphs a time series of the relative selection-corrected gender wage gaps estimated using individuals with a labor force participation rate of at least 80 percent. The square-marked series compares the gender wage gaps estimated using the "CPS-Similar" sample and the CPS. The figure shows that, between 1980 and 2010, the level of the gender wage gap is smaller in each year when using the "CPS-Similar sample, compared to when using the CPS. Namely, because the difference between the average male and female wage residuals (see Equation (1.3)) is smaller when using the "CPS-Similar sample, the sex regressor has less variability to explain.[29]

Additionally, how these differences in estimated wage gap levels evolve over time affects the narrowing of the gender wage gap. Figure 1.5 depicts the convergence of the selection-corrected wage gaps at each threshold between 1980 and 2010 for the CPS and "CPS-Similar" samples.[30] The 1980 Census (or 1975-1984 CPS) is used to control for both male and female selection into the full-time full-year labor force. In Figure 1.5, at each threshold, the shift to the "CPS-Similar" sample results in relatively less convergence. The largest difference in convergence is at the 50 percent threshold. In 1980, the average wage residual for women is larger in the CPS, relative to the "CPS-Similar" sample. By 2010, the difference between women's average residual in the CPS and women's average residual in the "CPS-Similar" sample becomes smaller. Because men's average wage residual in both samples are of similar magnitude and change little over time, the spread between men and womens average residuals falls less in the "CPS-Similar" sample relative to the CPS; the estimated gender wage gap narrows by 0.055 log points using the "CPS-Similar" sample and by 0.074 log points using CPS data.[31] Therefore, differences in wage

[29]Identification at infinity uses the full wage sample to estimate the OLS log hourly wage, $\mathbf{X}_{it}\hat{\beta}_t^{\infty}$, such that the vector of coefficients on $\mathbf{X}_{it}$ are the same regardless of sex or predicted labor force participation. Note that the predicted wage for females is the wage they would have received if they were males. Equation (1.3) then takes the difference between the observed and predicted OLS wage, and regresses this residual on sex.

[30]I calculate the convergence of the gender wage gap starting with 1980 rather than 1970 because the 1970 full Census sample only yields 16 women with a predicted full-time full-year labor force participation over 80 percent.

[31]Using the "CPS-Similar" sample, the difference between the male and female average wage residuals is 0.265 log points in 1980 and .21 log points in 2010. Therefore, the spread falls by 0.055 log points. Using the CPS, the spread decreases by 0.074 log points.

gap convergence at the 50 percent threshold are likely due to the use of different datasets. On the other hand, the gender wage gaps at the 80 percent threshold, estimated using the CPS and "CPS-Similar" samples, both converge by about 0.052 log points, despite the level of the gender wage gap at each point in time being relatively smaller using the "CPS-Similar" sample. As a result, at higher thresholds, differences in convergence between the CPS and full Census samples are not solely due to the use of different datasets.

Figure 1.6 compares the estimated convergence in the gender wage gap across three iterations of the Census sample: (1) the "CPS-Similar" sample using the 1980 Census to model male and female selection, (2) the full Census sample using the 1980 Census to model male and female selection, and (3) the full Census sample using the 1980 Census to model female selection and the 2010 ACS to model male selection. To illustrate the effects of including the group quarters population in the first and second stages of identification at infinity, I compare the selection-corrected gender wage gaps estimated using the "CPS-Similar sample and the full Census sample. For both samples, I use the 1980 Census to model male and female selection into the full-time full-year labor force. The largest difference in estimated convergence across these samples is at the 80 percent threshold, not because I expand the population used to estimate labor force participation, but because the selection of women with strong predicted labor force attachment appears to be sensitive to small changes in the first-stage probit coefficients in Table 1.C.7.[32] About 355 women in the 1980 "CPS-Similar" sample have a predicted probability of full-time full-year labor force participation above 80 percent, while in the full 1980 Census sample, only 98 women have a predicted probability exceeding 80 percent.[33] In 1980, because of the differences in the types of women that have a predicted labor force participation rate above 80 percent (see Table 1.C.11), the average wage residual among women who satisfy the 80

[32]At lower thresholds, the shift between the "CPS-Similar" and full Census samples only results in small changes in convergence.

[33]In the 1980 "CPS-Similar" sample, the maximum predicted probability of full-time full-year labor force participation among women in 81.9 percent. For men, the maximum predicted probability is 88.7 percent. In the 1980 full Census sample, the maximum predicted probability of full-time full-year labor force participation among women in 80.5 percent.

percent threshold is larger when using the full Census. But, by 2010, the difference in women's average wage residuals across the "CPS-Similar" and full Census samples decreases (average wage residuals for men do not differ much across time and samples). As a result, at the 80 percent threshold, the average wage residual among women in the "CPS-Similar" sample converges less than the average wage residual among women in the full Census.

Lastly, I show how accounting for declining male labor force participation by using the full 2010 ACS, rather than the full 1980 Census, to estimate the first-stage selection coefficients affects the level and convergence of the selection-corrected gender wage gap. For both samples, I use the 1980 Census to model female selection into the labor force. Therefore, any differences between the estimated gender wage gaps are due to changes in the sample of men selected for the second stage of identification at infinity. The triangle-marked series in Figure 1.4 shows that at the 80 percent threshold, using the 2010 ACS to estimate male selection yields a relatively smaller estimated gender wage gap at each point in time; the difference between male and female average wage residuals is smaller in each year. Furthermore, in Figure 1.6, the estimated gender wage gap narrows less when using the 2010 ACS to model male selection because the difference between male and female average wage residuals converges relatively less. In particular, by modifying how male selection into the labor force is modeled, the convergence in the selection-corrected wage gap falls from 0.08 log points to 0.061 log points.

Differences in convergence may be a result of how the type of man that is predicted to have strong labor force attachment changes when the 2010 ACS is used for the first stage of identification at infinity. Using either 1980 or 2010 data to model male selection into the full-time full-year labor force, men with a predicted probability of labor force participation greater than 80 percent ($P(\mathbf{X}_{i,\widehat{1980,2010}}\delta_{g=0}) \geq 80\%$), have, on average, about 16.6 years of education. Their average log hourly wage increased from 3.18 in 1980 to 3.27 in 2010. Men whose predicted probability is greater than 80 percent when using the 1980 Census to model male selection, but less than 80 percent when using the 2010 ACS ($P(\mathbf{X}_{i,\widehat{1980}\delta_{g=0}}) \geq 80\%$ and $P(\mathbf{X}_{i,\widehat{2010}\delta_{g=0}}) < 80\%$), have, on average, approximately 13 years of education. Their average log hourly wages fell from 3.01 in 1980 to 2.88 in 2010.[34] Therefore, the differences in convergence at the 80 percent threshold suggest that, in general, males are becoming more positively selected into the labor force and that not accounting for declining male labor force participation rates may overestimate the convergence gender wage gap.

In all, I extend on MR by estimating selection-corrected gender wage gaps using the Census to include individuals residing in group quarters in the first- and second-stage estimates. I also modify how selection is modeled for men to account for declining male labor force participation. I find that, relative to MR, using the full Census sample and the 2010 ACS to model male selection into the full-time full-year labor force yields relatively more convergence in the selection-corrected gender wage gaps. The x-marked series in Figure 1.4 shows that at the 80 percent threshold, differences between the CPS and full Census in the estimated wage gap level at each point in time are likely due to differences in the dataset used. Nonetheless, at the 80 percent threshold, changing how I model male selection appears to affect the convergence of the gender wage gap. Therefore, my results suggest that declining male labor force participation rates may be a factor in the narrowing of the gender wage gap, particularly among the most strongly attached men and women. Ultimately, my results using the Census and ACS data in an identification at infinity framework continue to support the results obtained by MR when using the CPS: under the assumption that the amount of selection bias decreases as the full-time full-year participation threshold increases, changing selection among women over time can help explain the convergence in the observed gender wage gap.

[34]Note, there are about 2,000 to 9,000 men in each time period whose predicted probability is less than 80 percent when using the 1980 Census to model male selection, but greater than 80 percent when using the 2010 ACS ($P(\widehat{\mathbf{X}_{i,1980}\delta_{g=0}}) < 80\%$ and $P(\widehat{\mathbf{X}_{i,2010}\delta_{g=0}}) \geq 80\%$). They have approximately 17.5 years of education, and their average log hourly wages increased from 2.99 in 1980 to 3.19 in 2010.

1.5 Conclusion

This paper replicates Mulligan and Rubinstein (2008) and extends on their work by using the Census to re-estimate selection-corrected gender wage gaps. To replicate MR's CPS results, I use the Heckman two-step method, with marital status interacted with the number of children under the age of 6 as the exclusion restriction, and identification at infinity to calculate a selection-corrected relative wage series for women between the 1970s and 1990s. There are small, but statistically significant, differences between the results in the replication and MR because I am unable to replicate the wage cutoffs and fixed and variable weights used in MR. Nonetheless, the Heckman two-step and identification at infinity results continue to suggest that after controlling for female selection into the full-time full-year labor force, much of the convergence in the gender wage gap disappears.

Yet, because the CPS excludes individuals residing in group quarters and selection is nonrandom, the first- and second-stage Heckman two-step estimates will be biased. The change in the gender wage gap over time will also be mis-estimated since the CPS is becoming more select over time due to the growing incarcerated population. Although the Census is more representative of the U.S. population and includes individuals residing in group quarters, it does not indicate how many children individuals in group quarters have. Therefore, I use the "CPS-Similar" Census sample, which excludes group quarters, to estimate the Heckman two-step model. I continue to find that changing female selection bias helps to explain the convergence of the gender wage gap.

Lastly, I extend the identification at infinity analysis to the Census because the select nature of the CPS makes it difficult to correctly identify demographic groups that have strong full-time full-year labor force participation rates and, thus, wages less subject to selection bias. The nonrandom omission of certain subpopulations will yield overestimated first-stage labor force participation rates and mis-estimated wage levels, particularly for they type of men who are more likely to reside in group quarters. The increasingly select nature of the CPS also mis-estimates the change in the gender wage gap because in later years there are demographic groups who may be subject to more selection bias than is evident in the CPS. Furthermore, unlike MR which uses 1975-1979 CPS data to select males with high labor force attachment, I use the 2010 ACS in the first stage of identification at infinity because fewer men worked full-time full-year in the 2000s. As a result of declining male labor force participation, using 1975-1979 data to select strongly attached men would yield a gender wage gap more likely to be subject to changing selection bias among men.

The differences between the identification at infinity results from the Census and the CPS appear to be driven by both the use of different datasets and the use of different selection rules which account for declining male labor force participation. I find that, similar to MR, the selection-corrected gender wage gap between strongly attached men and women did not close as much as the observed wage gap. However, there is more convergence in the gender wage gap estimated using the Census rather than the CPS. Overall, the identification at infinity results suggest that changing selection into the female and male full-time full-year labor force contributes to the narrowing of the observed gender wage gap.

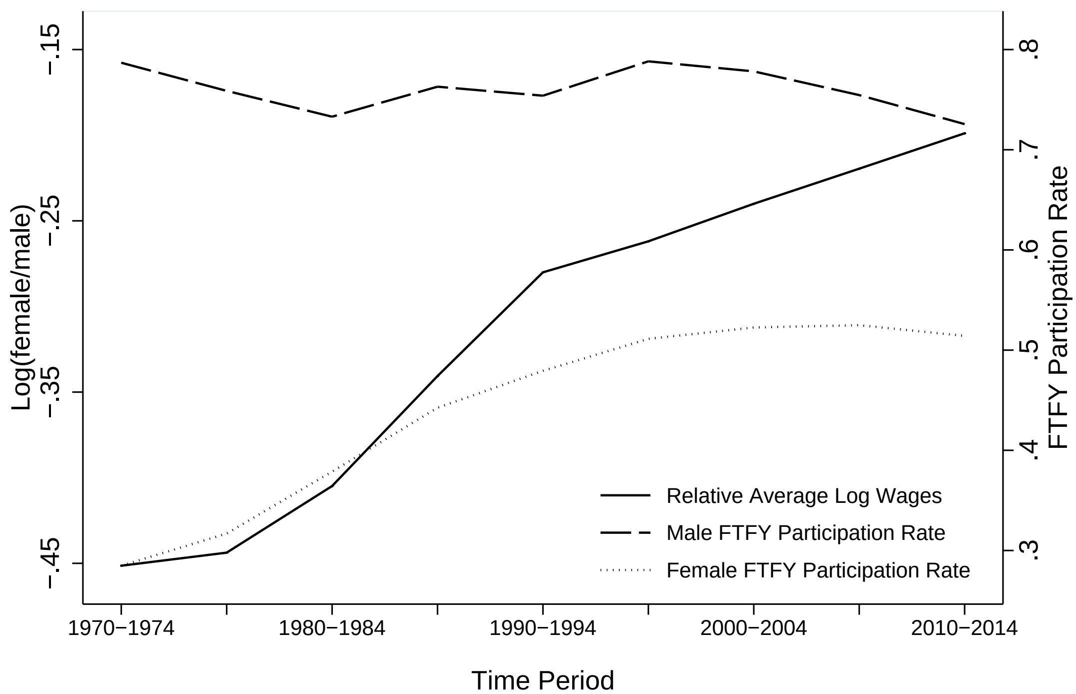


Figure 1.1: Gender wage gap and FTFY participation time series using the CPS

Using the CPS, the figure compares how the average log hourly wage of women relative to the average log hourly wage of men has changed over time (left y-axis). It also displays a time series of the male and female full-time full-year labor force participation rates (right y-axis). The sample consists of white, non-Hispanic individuals aged 26 to 55, not living in group quarters. Each time period is created by grouping 5 consecutive years of CPS ASEC data.

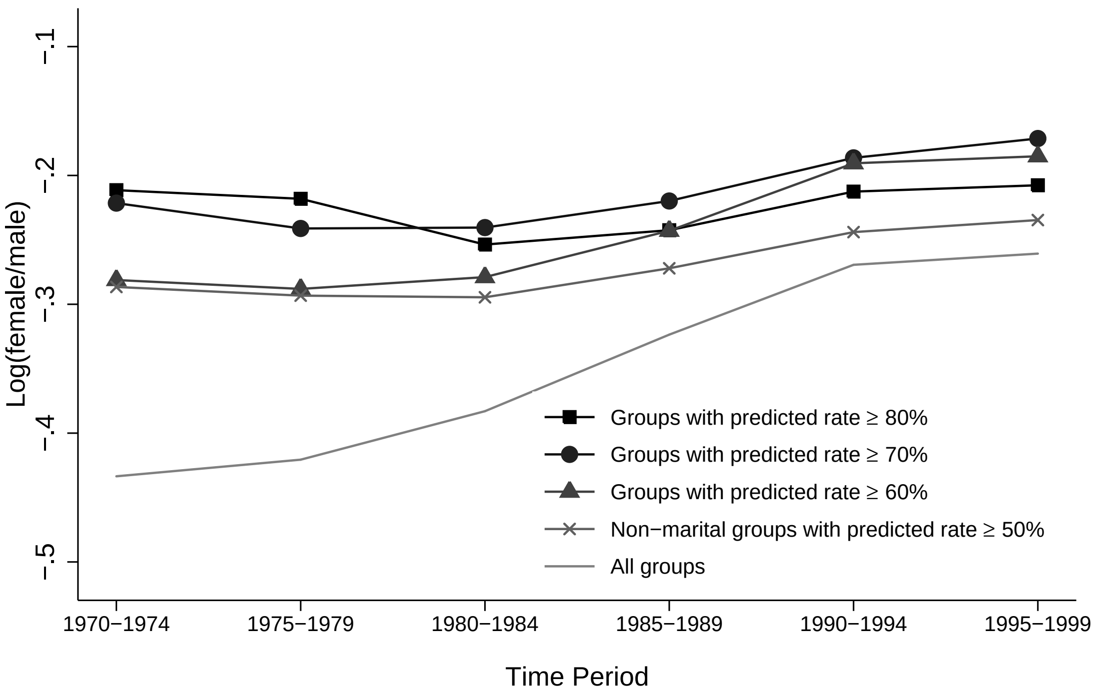


Figure 1.2: Gender wage gaps among strongly attached groups, various thresholds, using the CPS

Following Mulligan and Rubinstein (2008), the figure graphs five time series of women's log wages relative to men's, net of demographic characteristics. The series differ according to the demographic groups (defined according to gender, schooling, marital status, and potential experience, except for the x-marked series which does not use marital status) included in the estimation. The unmarked series includes all demographic groups to calculate the wage ratio over time. For all of the marked series, demographic groups are selected based on their full-time full-year labor force participation rate, which is predicted using the first-stage coefficients estimated with 1975 to 1979 (work-year) CPS data. The calculations use the CPS wage sample of white, non-Hispanic individuals aged 26 to 55. The wage trimming process is described in Section 1.3.1.

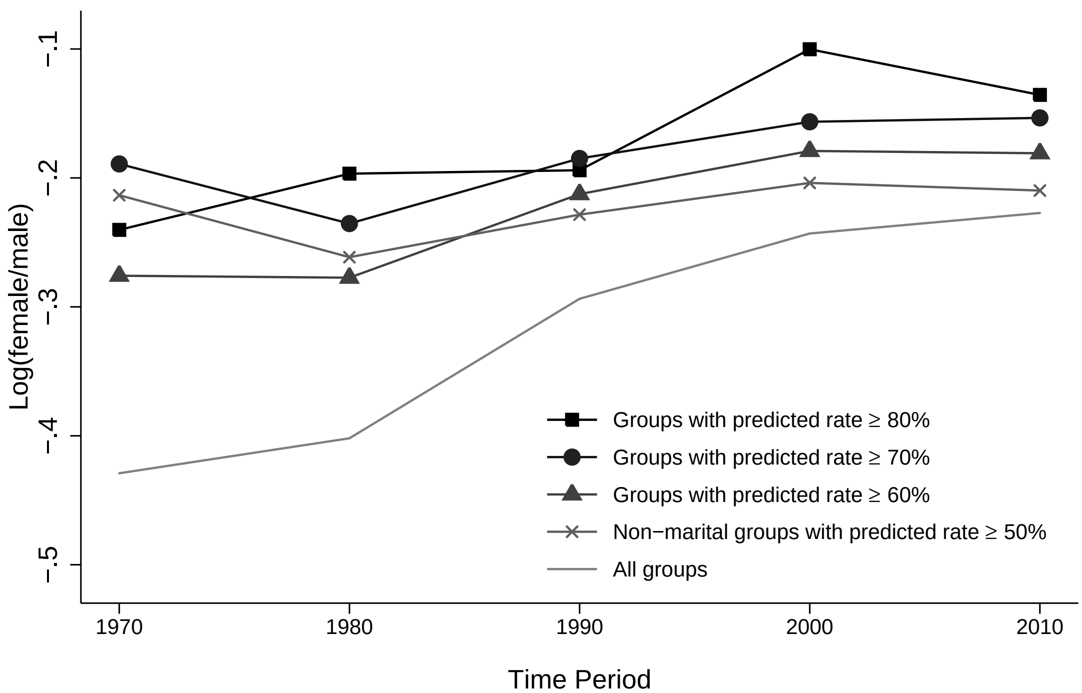


Figure 1.3: Gender wage gaps among strongly attached groups, various thresholds - Using the full Census sample and 1980 to model female selection and 2010 to model male selection

The figure graphs five time series of women's log wages relative to men's, net of demographic characteristics. The series differ according to the demographic groups (defined according to gender, schooling, marital status, and potential experience, except for the x-marked series which does not use marital status) included in the estimation. The unmarked series includes all demographic groups to calculate the wage ratio over time. For all of the marked series, female and male demographic groups are selected based on their predicted full-time full-year labor force participation rate. The 1980 Census data is used to estimate the first-stage female selection equation. The 2010 ACS is used to estimate the male selection equation. The calculations use the full Census wage sample of white, non-Hispanic individuals aged 26 to 55. Wage cutoffs are generated using Census data, following the wage trimming process described in Mulligan and Rubinstein (2008). Following Autor, Katz, and Kearney (2005), I impute hours and weeks worked for the 1970 Census and weeks worked for the 2010 ACS.

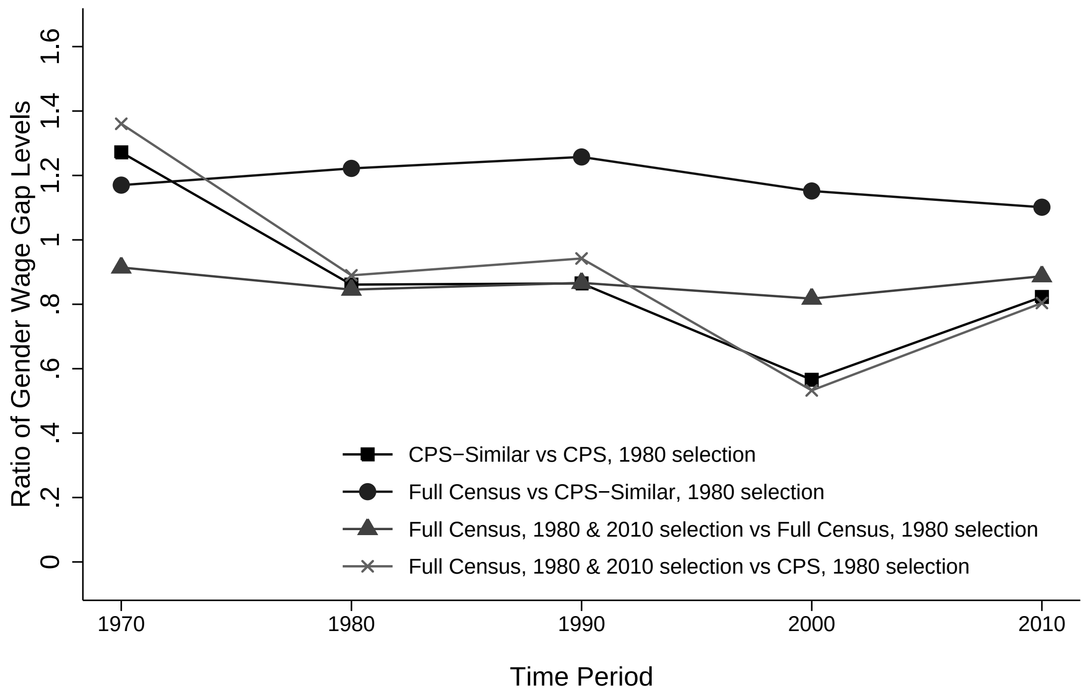


Figure 1.4: Relative gender wage gaps among strongly attached groups, 80% threshold

The figure graphs four time series of relative estimated gender wage gaps by comparing the following datasets: the CPS, the "CPS-Similar" sample, and the full Census sample. All samples use the 1980 Census to predict full-time full-year labor force participation for females. The triangle- and x-marked series use the 2010 ACS to model male selection into the labor force; all others use the 1980 Census for men's first-stage estimates. Each time series is calculated by taking the ratio of estimated gender wage gaps (using the 80% threshold for wage sample selection) at each point in time. For example, the square-marked series divides the "CPS-Similar" estimated wage gap by the CPS estimated wage gap, or $\hat{\gamma}_{t,CPS-Sim}$ divided by $\hat{\gamma}_{t,CPS}$ (see Equation (1.3)). The calculations use the wage samples of white, non-Hispanic individuals aged 26 to 55. Wage cutoffs are generated following the wage trimming process described in Mulligan and Rubinstein (2008).

Figure 1.5: Convergence of gender wage gaps among strongly attached groups, various thresholds - Comparing the CPS and "CPS-Similar" samples

The first stage of identification at infinity selects samples of demographic groups (defined according to gender, schooling, marital status, and potential experience, except for the 50 percent threshold which does not use marital status) such that their predicted probability of full-time full-year labor force participation exceeds some α near 1. There are two bars, at each level of alpha, representing the dataset used to estimate the selection-corrected wage gaps; the figure compares the CPS and the "CPS-Similar" sample. All samples use the 1975 to 1984 (work-year) pooled CPS or the 1980 Census to predict full-time full-year labor force participation for males and females. For each dataset and threshold, convergence is calculated by subtracting $\hat{\gamma}_{1980}$ from $\hat{\gamma}_{2010}$ (see Equation (1.3)), or the 1980 (1975-1984) estimated gender wage gap from the 2010 (2005-2014) gender wage gap. The sample consists of white, non-Hispanic individuals aged 26 to 55. Wage cutoffs are generated following the wage trimming process described in Mulligan and Rubinstein (2008).

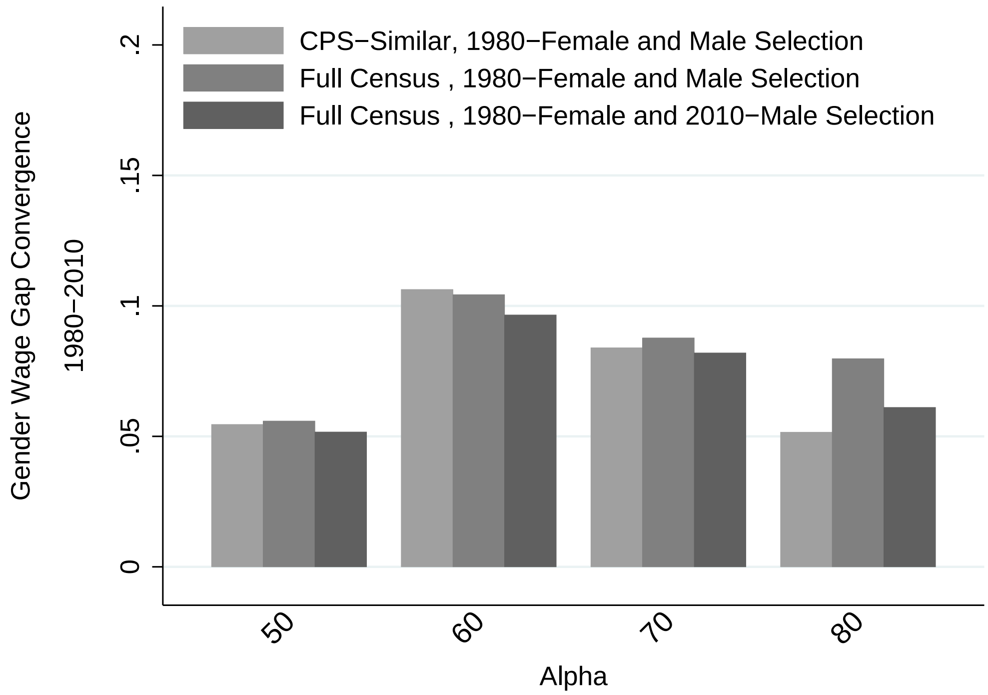


Figure 1.6: Convergence of gender wage gaps among strongly attached groups, various thresholds - Comparing the "CPS-Similar" and full Census samples

The first stage of identification at infinity selects samples of demographic groups (defined according to gender, schooling, marital status, and potential experience, except for the 50 percent threshold which does not use marital status) such that their predicted probability of full-time full-year labor force participation exceeds some α near 1. There are three bars, at each level of alpha, representing the dataset used to estimate the selection-corrected wage gaps. The figure compares the "CPS-Similar" sample and the full Census sample. All samples use the 1980 Census to predict full-time full-year labor force participation for females. One instance (the darkest bar) uses the 2010 ACS to model male selection into the labor force; all others use the 1980 Census for men's first-stage estimates. For each dataset and threshold, convergence is calculated by subtracting $\hat{\gamma}_{1980}$ from $\hat{\gamma}_{2010}$ (see Equation (1.3)), or the 1980 estimated gender wage gap from the 2010 gender wage gap. The sample consists of white, non-Hispanic individuals aged 26 to 55. Wage cutoffs are generated using Census data, following the wage trimming process described in Mulligan and Rubinstein (2008).

Table 1.1: Descriptive statistics for the full CPS sample

	MR (2008)		Replication		Extension	
	1975-1979 Female (1)	1995-1999 Female (2)	1975-1979 Female (3)	1995-1999 Female (4)	1975-1979 Male (5)	1995-1999 Male (6)
Total Observations	116,843	102,395	116,843	102,395	112,291	99,033
FTFY Indicator	0.317	0.511	0.317	0.511	0.759	0.788
Widowed	0.031	0.018	0.031	0.018	0.006	0.005
Divorced	0.085	0.138	0.085	0.138	0.060	0.114
Separated	0.023	0.025	0.023	0.025	0.017	0.020
Never Married	0.063	0.118	0.063	0.118	0.107	0.184
0-8 school years	0.071	0.015	0.070	0.015	0.090	0.018
9-11 school years	0.135	0.048	0.135	0.048	0.117	0.055
High School Degree	0.451	0.345	0.451	0.345	0.330	0.332
College Degree	0.127	0.209	0.127	0.209	0.170	0.212
Advanced Degree	0.032	0.089	0.032	0.089	0.088	0.109
Potential Experience - 15	5.073	4.527	5.067	4.527	4.396	4.335
Midwest	0.284	0.267	0.284	0.267	0.282	0.269
South	0.303	0.330	0.303	0.330	0.303	0.326
West	0.176	0.197	0.176	0.197	0.182	0.201
Kids 0-6	0.354	0.324	0.354	0.324	0.390	0.318

The full sample drawn from the CPS consists of white, non-Hispanic individuals between the ages of 26 and 55, not residing in group quarters. Omitted groups are currently married, some college, 15 years' potential experience, and Northeast region. The first two columns are taken directly from Mulligan and Rubinstein (2008). Columns 3 and 4 list the replication results. Columns 5 and 6 list the descriptive statistics for all males in the sample.

Table 1.2: Descriptive statistics for the FTFY CPS sample

	MR (2008)				Replication			
	1975-1979 Female (1)	1995-1999 Female (2)	1975-1979 Male (3)	1995-1999 Male (4)	1975-1979 Female (5)	1995-1999 Female (6)	1975-1979 Male (7)	1995-1999 Male (8)
Total Observations	27,656	41,062	57,457	54,565	27,578	41,004	58,003	55,015
FTFY Indicator	1	1	1	1	1	1	1	1
Log Real Hourly Wage	2.487	2.607	2.924	2.861	2.492	2.615	2.936	2.877
Inverse Mills	0.997	0.727	0	0	0.995	0.727	0	0
Widowed	0.037	0.017	0.005	0.004	0.037	0.017	0.005	0.004
Divorced	0.158	0.178	0.048	0.104	0.158	0.177	0.048	0.103
Separated	0.028	0.026	0.014	0.018	0.028	0.026	0.014	0.018
Never Married	0.128	0.151	0.082	0.156	0.128	0.151	0.082	0.155
0-8 school years	0.042	0.005	0.066	0.010	0.041	0.005	0.065	0.009
9-11 school years	0.096	0.030	0.107	0.042	0.096	0.029	0.106	0.042
High School Degree	0.444	0.327	0.342	0.327	0.443	0.326	0.340	0.324
College Degree	0.164	0.229	0.182	0.231	0.164	0.229	0.184	0.233
Advanced Degree	0.055	0.106	0.085	0.104	0.056	0.107	0.087	0.108
Potential Experience - 15	4.112	4.139	3.564	3.742	4.102	4.136	3.585	3.760
Midwest	0.278	0.282	0.293	0.289	0.278	0.282	0.292	0.288
South	0.334	0.343	0.301	0.331	0.333	0.342	0.301	0.331
West	0.170	0.182	0.169	0.188	0.170	0.182	0.169	0.188
Kids 0-6	0.158	0.207	0.446	0.351	0.158	0.208	0.444	0.352

The FTFY sample drawn from the CPS consists of full-time full-year white, non-Hispanic individuals between the ages of 26 and 55, not residing in group quarters. Further sample restrictions, including wage trimming, are described in the text. Omitted groups are currently married, some college, 15 years' potential experience, and Northeast region. The first four columns are taken directly from Mulligan and Rubinstein (2008). Columns 5 through 8 are a replication of MR using their CPS data and methodology.

Table 1.3: Correcting the gender wage gap using the Heckman two-step estimator - Using the CPS to control for female selection

	MR (2008)			Replication		
	Method			Method		
Period	OLS (1)	Two-Step (2)	Bias (3)	OLS (4)	Two-Step (5)	Bias (6)
Panel A: Variable Weights						
1975-1979	-0.414	-0.337	-0.077	-0.417	-0.339	-0.078
	(0.003)	(0.014)	(0.015)	(0.003)	(0.014)	(0.014)
1995-1999	-0.254	-0.339	0.085	-0.259	-0.350	0.091
	(0.003)	(0.014)	(0.015)	(0.003)	(0.014)	(0.015)
Change	0.160	-0.002	0.162	0.158	-0.011	0.169
	(0.005)	(0.020)	(0.021)	(0.004)	(0.020)	(0.020)
Panel B: Fixed Weights						
1975-1979	-0.404	-0.330	-0.075	-0.407	-0.333	-0.074
	(0.003)	(0.014)	(0.014)	(0.003)	(0.013)	(0.014)
1995-1999	-0.264	-0.353	0.089	-0.264	-0.358	0.094
	(0.004)	(0.015)	(0.016)	(0.003)	(0.015)	(0.015)
Change	0.140	-0.024	0.164	0.143	-0.025	0.168
	(0.005)	(0.021)	(0.021)	(0.005)	(0.020)	(0.020)

As in Mulligan and Rubinstein (2008), each cell summarizes the regression results listed in Tables 1.C.1, 1.C.2, and 1.C.3. The first three columns are taken directly from Mulligan and Rubinstein (2008). The last three columns list the replication results. The entries are female minus male log wages, which differ from each other in terms of (a) rows, or time period used for estimation (1975-1979 vs 1995-1999); (b) columns, or whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or whether variable or fixed weights are used to calculate log wages (see Appendix 1.A). The "Bias" column is the difference between the OLS and two-step columns. The "Change" row is the difference between the 1995-1999 and 1975-1979 rows.

The regressions control for demographics interacted with sex and use the CPS wage sample of white, non-Hispanic individuals aged 26 to 55. The wage trimming process is described in Section 1.3.1. Bootstrap standard errors are in parentheses.

Table 1.4: Gender-gap changes by marital status and schooling - Using the CPS to control for female selection

	MR (2008)					Replication				
	OLS			Two-Step	Bias	OLS			Two-Step	Bias
	1975-1979 (1)	1995-1999 (2)	Change (3)	Change (4)	Change (5)	1975-1979 (6)	1995-1999 (7)	Change (8)	Change (9)	Change (10)
Panel A: All										
Conditional on marital status	-0.404	-0.264	0.140	-0.024	0.164	-0.407	-0.264	0.143	-0.025	0.168
	(0.003)	(0.004)	(0.005)	(0.021)	(0.021)	(0.003)	(0.003)	(0.005)	(0.020)	(0.020)
Not conditional on marital status	-0.431	-0.270	0.160	-0.009	0.169	-0.436	-0.273	0.164	-0.006	0.170
	(0.003)	(0.004)	(0.005)	(0.017)	(0.018)	(0.003)	(0.003)	(0.004)	(0.013)	(0.014)
Panel B: By Marital Status										
Currently Married	-0.471	-0.311	0.161	-0.026	0.187	-0.477	-0.313	0.164	-0.028	0.193
	(0.004)	(0.004)	(0.005)	(0.023)	(0.024)	(0.004)	(0.004)	(0.005)	(0.023)	(0.023)
Separated	-0.380	-0.293	0.087	-0.066	0.153	-0.379	-0.296	0.083	-0.073	0.157
	(0.021)	(0.021)	(0.030)	(0.035)	(0.046)	(0.021)	(0.022)	(0.030)	(0.035)	(0.046)
Widowed	-0.430	-0.252	0.178	0.019	0.159	-0.432	-0.251	0.180	0.017	0.164
	(0.025)	(0.042)	(0.049)	(0.053)	(0.072)	(0.025)	(0.042)	(0.049)	(0.052)	(0.071)
Divorced	-0.326	-0.189	0.136	0.019	0.117	-0.329	-0.190	0.139	0.019	0.120
	(0.010)	(0.009)	(0.013)	(0.020)	(0.024)	(0.010)	(0.008)	(0.013)	(0.019)	(0.023)
Never Married	-0.179	-0.127	0.052	-0.062	0.114	-0.185	-0.127	0.058	-0.058	0.117
	(0.010)	(0.009)	(0.013)	(0.019)	(0.023)	(0.009)	(0.009)	(0.013)	(0.018)	(0.022)
Panel C: By Education										
0 to 8 years	-0.378	-0.322	0.056	-0.206	0.262	-0.381	-0.349	0.032	-0.241	0.273
	(0.035)	(0.091)	(0.098)	(0.103)	(0.142)	(0.035)	(0.087)	(0.094)	(0.099)	(0.136)
High School, not grad	-0.429	-0.243	0.185	-0.373	0.222	-0.430	-0.239	0.190	-0.041	0.232
	(0.018)	(0.032)	(0.037)	(0.046)	(0.059)	(0.018)	(0.032)	(0.037)	(0.047)	(0.059)
High school graduates	-0.427	-0.297	0.130	-0.037	0.167	-0.425	-0.294	0.131	-0.042	0.173
	(0.007)	(0.009)	(0.011)	(0.023)	(0.026)	(0.007)	(0.008)	(0.011)	(0.023)	(0.025)
Some College	-0.409	-0.258	0.151	-0.008	0.159	-0.410	-0.256	0.154	-0.012	0.165
	(0.010)	(0.010)	(0.014)	(0.024)	(0.028)	(0.010)	(0.009)	(0.013)	(0.023)	(0.027)
College	-0.400	-0.237	0.163	0.012	0.151	-0.413	-0.245	0.168	0.011	0.157
	(0.013)	(0.011)	(0.017)	(0.025)	(0.030)	(0.013)	(0.012)	(0.017)	(0.025)	(0.031)
Advanced Degree	-0.276	-0.179	0.096	-0.018	0.115	-0.306	-0.203	0.103	-0.016	0.120
	(0.023)	(0.017)	(0.028)	(0.032)	(0.043)	(0.023)	(0.018)	(0.029)	(0.032)	(0.043)

Each cell summarizes the regression results listed in Tables 1.C.1, 1.C.2, and 1.C.3. The first five columns are taken directly from Mulligan and Rubinstein (2008). Columns 6 through 10 list the replication results. The entries are female minus male log wages, which differ from each other in terms of (a) rows, or demographic groups; (b) columns, or the time period used for estimation and whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or the types of demographic groups summarized. Time-invariant, or fixed, weights are used to calculate log wages. The "Bias" column is the difference between the OLS and two-step columns. Bootstrap standard errors are in parentheses.

Table 1.5: Descriptive statistics for the 1980 Census

	CPS-Similar Sample				Full Census Sample			
	Full Sample		FTFY Sample		Full Sample		FTFY Sample	
	Female (1)	Male (2)	Female (3)	Male (4)	Female (5)	Male (6)	Female (7)	Male (8)
Total Observations	1,705,976	1,652,348	433,145	879,403	1,716,016	1,672,448	433,839	880,536
Original FTFY Indicator	0.313	0.722			0.313	0.729		
Modified FTFY Indicator	0.337	0.734			0.337	0.741		
Log Real Hourly Wage			2.504	2.966			2.503	2.965
Inverse Mills			1.026	0			1.029	0
Widowed	0.028	0.005	0.031	0.004	0.028	0.005	0.031	0.004
Divorced	0.099	0.075	0.173	0.063	0.099	0.076	0.173	0.064
Separated	0.022	0.018	0.027	0.015	0.022	0.018	0.027	0.015
Never Married	0.074	0.112	0.135	0.087	0.076	0.116	0.136	0.088
0-8 school years	0.062	0.082	0.036	0.058	0.064	0.084	0.036	0.058
9-11 school years	0.133	0.114	0.101	0.100	0.133	0.114	0.101	0.100
High School Degree	0.422	0.324	0.444	0.337	0.420	0.323	0.444	0.337
College Degree	0.129	0.168	0.127	0.186	0.129	0.167	0.127	0.186
Advanced Degree	0.039	0.096	0.046	0.087	0.040	0.095	0.047	0.087
Potential Experience - 15	4.525	3.885	3.905	3.495	4.535	3.886	3.904	3.494
Midwest	0.278	0.278	0.269	0.289	0.278	0.278	0.269	0.289
South	0.313	0.313	0.331	0.308	0.312	0.314	0.331	0.308
West	0.181	0.186	0.174	0.168	0.181	0.187	0.174	0.168
Kids 0-6	0.357	0.392	0.163	0.416	0.355	0.388	0.163	0.416
Institutionalized GQ					0.003	0.007	0.0002	0.0004
Non-Institutionalized GQ					0.002	0.005	0.0009	0.0009

The table lists descriptive statistics using the 1980 Census. The sample consists of white, non-Hispanic individuals between the ages of 26 and 55. The “CPS-Similar” sample, drawn from the Census, does not include individuals residing in group quarters. The full Census sample includes individuals living in group quarters. Following Mulligan and Rubinstein (2008), omitted groups are currently married, some college, 15 years’ potential experience, and Northeast region. Columns 1, 2, 5, and 6 list the descriptive statistics for the entire sample. Columns 3, 4, 7, and 8 list the descriptive statistics for the full-time full-year sample (using the original definition of full-time full-year). Wage cutoffs are generated using Census data, following the wage trimming process described in Mulligan and Rubinstein (2008).

Table 1.6: Descriptive statistics for the 2000 Census

	CPS-Similar Sample				Full Census Sample			
	Full Sample		FTFY Sample		Full Sample		FTFY Sample	
	Female (1)	Male (2)	Female (3)	Male (4)	Female (5)	Male (6)	Female (7)	Male (8)
Total Observations	2,162,273	2,100,306	734,504	1,011,829	2,175,263	2,139,521	734,988	1,013,766
Original FTFY Indicator	0.477	0.749			0.475	0.738		
Modified FTFY Indicator	0.505	0.758			0.503	0.746		
Log Real Hourly Wage			2.674	2.924			2.674	2.923
Inverse Mills			0.783	0			0.787	0
Widowed	0.017	0.005	0.015	0.003	0.017	0.005	0.015	0.003
Divorced	0.145	0.119	0.186	0.105	0.145	0.120	0.186	0.105
Separated	0.023	0.017	0.022	0.015	0.023	0.018	0.022	0.015
Never Married	0.123	0.182	0.160	0.153	0.124	0.184	0.160	0.153
0-8 school years	0.014	0.018	0.004	0.007	0.015	0.020	0.004	0.007
9-11 school years	0.045	0.053	0.023	0.036	0.045	0.055	0.023	0.036
High School Degree	0.299	0.306	0.266	0.278	0.299	0.308	0.266	0.278
College Degree	0.204	0.206	0.225	0.236	0.203	0.203	0.225	0.235
Advanced Degree	0.102	0.113	0.114	0.116	0.102	0.111	0.114	0.116
Potential Experience - 15	4.919	4.856	4.436	4.208	4.935	4.865	4.434	4.203
Midwest	0.265	0.266	0.281	0.285	0.265	0.265	0.281	0.285
South	0.338	0.337	0.339	0.330	0.338	0.338	0.339	0.330
West	0.192	0.196	0.177	0.183	0.192	0.197	0.177	0.183
Kids 0-6	0.343	0.340	0.226	0.371	0.341	0.334	0.225	0.370
Institutionalized GQ					0.002	0.013	0.0002	0.0011
Non-Institutionalized GQ					0.003	0.006	0.0004	0.0007

The table lists descriptive statistics using the 2000 Census. The sample consists of white, non-Hispanic individuals between the ages of 26 and 55. The "CPS-Similar" sample, drawn from the Census, does not include individuals residing in group quarters. The full Census sample includes individuals living in group quarters. Following Mulligan and Rubinstein (2008), omitted groups are currently married, some college, 15 years' potential experience, and Northeast region. Columns 1, 2, 5, and 6 list the descriptive statistics for the entire sample. Columns 3, 4, 7, and 8 list the descriptive statistics for the full-time full-year sample (using the original definition of full-time full-year). Wage cutoffs are generated using Census data, following the wage trimming process described in Mulligan and Rubinstein (2008).

Table 1.7: Descriptive statistics for the 2010 ACS

	CPS-Similar Sample				Full Census Sample			
	Full Sample		FTFY Sample		Full Sample		FTFY Sample	
	Female (1)	Male (2)	Female (3)	Male (4)	Female (5)	Male (6)	Female (7)	Male (8)
Total Observations	416,743	398,870	164,622	199,219	418,997	407,486	164,668	199,153
Original FTFY Indicator	0.501	0.687			0.498	0.676		
Modified FTFY Indicator	0.513	0.691			0.510	0.679		
Log Real Hourly Wage			2.728	2.930			2.728	2.931
Inverse Mills			0.751	0			0.754	0
Widowed	0.016	0.005	0.013	0.004	0.016	0.006	0.013	0.004
Divorced	0.150	0.126	0.166	0.103	0.151	0.128	0.166	0.103
Separated	0.028	0.020	0.023	0.016	0.028	0.020	0.023	0.016
Never Married	0.167	0.237	0.189	0.189	0.169	0.243	0.189	0.189
0-8 school years	0.012	0.015	0.003	0.005	0.013	0.017	0.003	0.005
9-11 school years	0.035	0.047	0.014	0.025	0.035	0.049	0.014	0.025
High School Degree	0.251	0.305	0.205	0.259	0.252	0.308	0.205	0.259
College Degree	0.237	0.218	0.273	0.262	0.236	0.214	0.273	0.262
Advanced Degree	0.126	0.111	0.164	0.132	0.125	0.109	0.164	0.132
Potential Experience - 15	5.578	5.772	4.881	5.044	5.591	5.778	4.881	5.045
Midwest	0.264	0.265	0.281	0.279	0.264	0.265	0.281	0.279
South	0.348	0.344	0.346	0.340	0.348	0.345	0.346	0.340
West	0.193	0.198	0.176	0.188	0.193	0.199	0.177	0.188
Kids 0-6	0.343	0.326	0.254	0.365	0.341	0.319	0.254	0.365
Institutionalized GQ					0.003	0.014	0.0001	0.0005
Non-Institutionalized GQ					0.003	0.006	0.0002	0.0006

The table lists descriptive statistics using the 2010 ACS. The sample consists of white, non-Hispanic individuals between the ages of 26 and 55. The “CPS-Similar” sample, drawn from the ACS, does not include individuals residing in group quarters. The full ACS sample includes individuals living in group quarters. Following Mulligan and Rubinstein (2008), omitted groups are currently married, some college, 15 years’ potential experience, and Northeast region. Columns 1, 2, 5, and 6 list the descriptive statistics for the entire sample. Columns 3, 4, 7, and 8 list the descriptive statistics for the full-time full-year sample (using the original definition of full-time full-year). Wage cutoffs are generated using ACS data, following the wage trimming process described in Mulligan and Rubinstein (2008).

Table 1.8: Correcting the gender wage gap using the Heckman two-step estimator - Using the "CPS-Similar" sample to control for female selection, 1980 v 2000

	Method		
Period	OLS (1)	Two-Step (2)	Bias (3)
	Panel A: Variable Weights		
1980	-0.424	-0.354	-0.069
	(0.002)	(0.004)	(0.005)
2000	-0.248	-0.340	0.092
	(0.001)	(0.004)	(0.005)
Change	0.176	0.014	0.162
	(0.002)	(0.006)	(0.007)
	Panel B: Fixed Weights		
1980	-0.420	-0.352	-0.068
	(0.001)	(0.004)	(0.005)
2000	-0.255	-0.351	0.096
	(0.002)	(0.005)	(0.005)
Change	0.165	0.001	0.164
	(0.002)	(0.006)	(0.007)

As in Mulligan and Rubinstein (2008), each cell summarizes the regression results listed in Tables 1.C.4 and 1.C.5. The entries are female minus male log wages, which differ from each other in terms of (a) rows, or time period used for estimation (1980 vs 2000); (b) columns, or whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or whether variable or fixed weights are used to calculate log wages (see Appendix 1.A). The "Bias" column is the difference between the OLS and two-step columns. The "Change" row is the difference between the 2000 and 1980 rows.

The regressions control for demographics interacted with sex and use the "CPS-Similar" wage sample of white, non-Hispanic individuals aged 26 to 55. Wage cutoffs are generated using Census data, following the wage trimming process described in Mulligan and Rubinstein (2008). Bootstrap standard errors are in parentheses.

Table 1.9: Correcting the gender wage gap using the Heckman two-step estimator - Using the "CPS-Similar" sample to control for female selection, 1980 v 2010

	Method		
Period	OLS (1)	Two-Step (2)	Bias (3)
	Panel A: Variable Weights		
1980	-0.424	-0.354	-0.069
	(0.002)	(0.004)	(0.005)
2000	-0.248	-0.340	0.092
	(0.001)	(0.004)	(0.005)
2010	-0.222	-0.321	0.099
	(0.003)	(0.013)	(0.014)
Change	0.201	0.033	0.168
	(0.003)	(0.014)	(0.014)
	Panel B: Fixed Weights		
1980	-0.420	-0.352	-0.068
	(0.001)	(0.004)	(0.005)
2000	-0.255	-0.351	0.096
	(0.002)	(0.005)	(0.005)
2010	-0.236	-0.344	0.108
	(0.004)	(0.015)	(0.015)
Change	0.184	0.008	0.176
	(0.004)	(0.015)	(0.016)

As in Mulligan and Rubinstein (2008), each cell summarizes the regression results listed in Tables 1.C.4 and 1.C.5. The entries are female minus male log wages, which differ from each other in terms of (a) rows, or time period used for estimation (1980, 2000, and 2010); (b) columns, or whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or whether variable or fixed weights are used to calculate log wages (see Appendix 1.A). The "Bias" column is the difference between the OLS and two-step columns. The "Change" row is the difference between the 2010 and 1980 rows.

The regressions control for demographics interacted with sex and use the "CPS-Similar" wage sample of white, non-Hispanic individuals aged 26 to 55. Wage cutoffs are generated using Census data, following the wage trimming process described in Mulligan and Rubinstein (2008). Bootstrap standard errors are in parentheses.

Table 1.10: Gender-gap changes by marital status and schooling - Using the "CPS-Similar" sample to control for female selection, 1980 v 2000

	OLS			Two-Step	Bias
	1980	2000	Change	Change	Change
	(1)	(2)	(3)	(4)	(5)
Panel A: All					
Conditional on marital status	-0.420	-0.255	0.165	0.001	0.164
	(0.001)	(0.002)	(0.002)	(0.006)	(0.007)
Not conditional on marital status	-0.446	-0.264	0.182	0.033	0.149
	(0.001)	(0.002)	(0.002)	(0.004)	(0.005)
Panel B: By Marital Status					
Currently Married	-0.487	-0.306	0.180	-0.005	0.185
	(0.001)	(0.002)	(0.002)	(0.007)	(0.007)
Separated	-0.423	-0.264	0.159	0.002	0.157
	(0.005)	(0.006)	(0.008)	(0.010)	(0.012)
Widowed	-0.411	-0.233	0.178	0.013	0.165
	(0.008)	(0.010)	(0.012)	(0.014)	(0.018)
Divorced	-0.359	-0.188	0.171	0.048	0.123
	(0.002)	(0.002)	(0.003)	(0.006)	(0.007)
Never Married	-0.208	-0.117	0.091	-0.033	0.125
	(0.002)	(0.002)	(0.003)	(0.006)	(0.007)
Panel C: By Education					
0 to 8 years	-0.363	-0.261	0.102	-0.154	0.255
	(0.016)	(0.051)	(0.053)	(0.054)	(0.076)
High School, not grad	-0.452	-0.327	0.125	-0.092	0.217
	(0.005)	(0.009)	(0.010)	(0.013)	(0.017)
High school graduates	-0.443	-0.281	0.162	0.000	0.161
	(0.001)	(0.002)	(0.002)	(0.006)	(0.007)
Some College	-0.411	-0.241	0.171	0.021	0.150
	(0.002)	(0.001)	(0.002)	(0.006)	(0.006)
College	-0.439	-0.238	0.201	0.031	0.170
	(0.004)	(0.003)	(0.005)	(0.008)	(0.009)
Advanced Degree	-0.314	-0.192	0.122	-0.042	0.164
	(0.011)	(0.006)	(0.012)	(0.014)	(0.018)

Using the "CPS-Similar" sample, each cell summarizes the regression results listed in Tables 1.C.4 and 1.C.5. The entries are female minus male log wages, which differ from each other in terms of (a) rows, or demographic groups; (b) columns, or the time period used for estimation (1980 and 2000) and whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or the types of demographic groups summarized. Time-invariant, or fixed, weights are used to calculate log wages. The "Bias" column is the difference between the OLS and two-step columns. Bootstrap standard errors are in parentheses.

Appendix

1.A Creating Fixed and Variable Weights

As in MR, tables that list the estimated gender wage gaps, such as Tables 1.3 and 1.4, use variable and fixed weights to estimate log hourly wages. Variable weights differ by time period, while fixed weights are the same across time periods. MR use identical values, except for sex, for the male and female weights to compare how the estimated gender wage gap is affected over time by changes in coefficients and not differences between the average full-time full-year male and female.

Following MR, to create variable weights I use the sample of full-time full-year women within each time period to calculate the average of the variables in the vector **Z**.[35] I generate fixed weights by first pooling all 10 years of CPS data, or the 1980 and 2000 Censuses, and then using the pooled sample of full-time full-year women to calculate the average of the demographic variables to create a weight that is the same across time periods.

To calculate the OLS gender wage gap, I substitute the weights into the OLS log hourly wage regressions and then take the difference between the estimated female and male log hourly wage. To calculate the Heckman two-step gender wage gap, I first use the probit coefficients and weights to estimate an inverse Mills' ratio for each female. Then I use the coefficients from the Heckman wage regression to estimate women's log hourly wages and take the difference between

[35] **Z** contains educational attainment, marital status, region, a potential experience quartic interacted with education indicators, and marital status interacted with the number of children under the age of 6.

the Heckman two-step-estimated female and OLS-estimated male log hourly wage to estimate the gender wage gap.

MR use fixed weights to ensure that the average full-time full-year woman does not change over time when estimating the gender wage gap. MR compare the estimated gender wage gaps using variable and fixed weights to examine whether results are affected by changes in the average full-time full-year woman over time.

Tables that estimate the "partial" effects of marital status and education, such as Table 1.4, use fixed weights. The weights in Panel B are calculated under the assumption that all full-time full-year women are of the marital status in a given row. All other averages of non-marital status variables are unchanged. Similarly, in Panel C, MR assume that all full-time full-year women attained a given level of education. Therefore, each category of marital status and educational attainment will take on a value of 0 or 1, in their respective panels, rather than a mean value between 0 and 1. Again, female and male weights are identical except for sex. Therefore, the male weights also take on a value of 0 or 1 for marital status in Panel B or for educational attainment in Panel C. As described above, I use the log hourly wage coefficients from the OLS and Heckman two-step regressions and the weights to estimate log hourly wages and then take the difference between the estimated female and male log wages to estimate the gender wage gap.

1.B Factor Variables

MR use pre-generated interactions in their Heckman two-step estimations; a separate variable for each interaction is included in the regressions. While using pre-generated interactions does not affect the probit, OLS, and Heckman two-step coefficients, it results in incorrect marginal effects and mis-estimated gender wage gaps.

As demonstrated in Williams (2012), if two variables and their interaction are included as separate variables, then when calculating the marginal effects of the interaction, Stata does not take into account that a change in the interaction affects the elements of the interaction individually. Consequently, for the Heckman two-step estimation, I use factor variables and re-estimate the marginal effects.

Table 1.C.12 lists the marginal effects from the first stage of the Heckman two-step method using the same methodology described in Section 1.3.2.1 but using factor variables. Inclusion of factor variables, rather than pre-generated interactions, does not appear to qualitatively change the results. The primary difference between the marginal effects from MR (see Table 1.C.1) and from the factor variable replication is the lack of a marginal effect reported for interactions in the replication; the value of the interaction term depends on the values of its component terms, so a separate marginal effect of the interaction cannot be estimated (Williams 2016).

I also re-estimate the gender wage gaps from Tables 1.3 and 1.4 using factor variables.[36] Although, pre-generated interactions do not affect the probit and wage coefficients, they affect the fixed and variable weights used to estimate log hourly wages and therefore the estimated gender wage gap. Rather than using the average of pre-generated interactions, I use the product of the average of the component variables when substituting the weights into the wage regressions. Tables 1.C.13 and 1.C.14 list the resulting estimated gender wage gaps. The differences between the gender wage gaps estimated using factor variables and the wage gaps estimated use pre-

[36]Following MR, standard errors are calculated using the non-parametric pairwise bootstrap method (1,000 replications) with seeds of 123, 456, 78, and 91.

generated interactions are statistically significant but qualitatively similar. Within Panel A, comparing Column 1 of Table 1.C.13 and Column 4 of Table 1.3, the OLS-estimated gender wage gap converges less when using factor variables (0.152 log points relative to 0.158 log points). However, relative to Column 6 of Table 1.3, selection bias is larger in Column 3 of Table 1.C.13. For example, in Panel A, selection bias, calculated using pre-generated interactions, changes by 0.169 log points between the 1970s and 1990s. The change in selection bias calculated using factor variables is 0.171 log points. Despite a reduced OLS-estimated wage gap convergence when using factor variables, the difference in the changes in selection bias can be attributed to the larger change in the Heckman two-step-estimated wage gap. In Column 5 of Table 1.3, the selection-corrected wage gap increases by 0.011 log points, while in Column 2 of Table 1.C.13, the estimated wage gap increases by 0.019 log points (note both of these values are statistically insignificant, but their difference is signficant). Table 1.C.14 tells a similar story. Overall, Tables 1.C.13 and 1.C.14 continue to illustrate that the OLS the gender wage gap converges over time, but once selection is controlled for, convergence disappears.

1.C Additional Tables and Figures

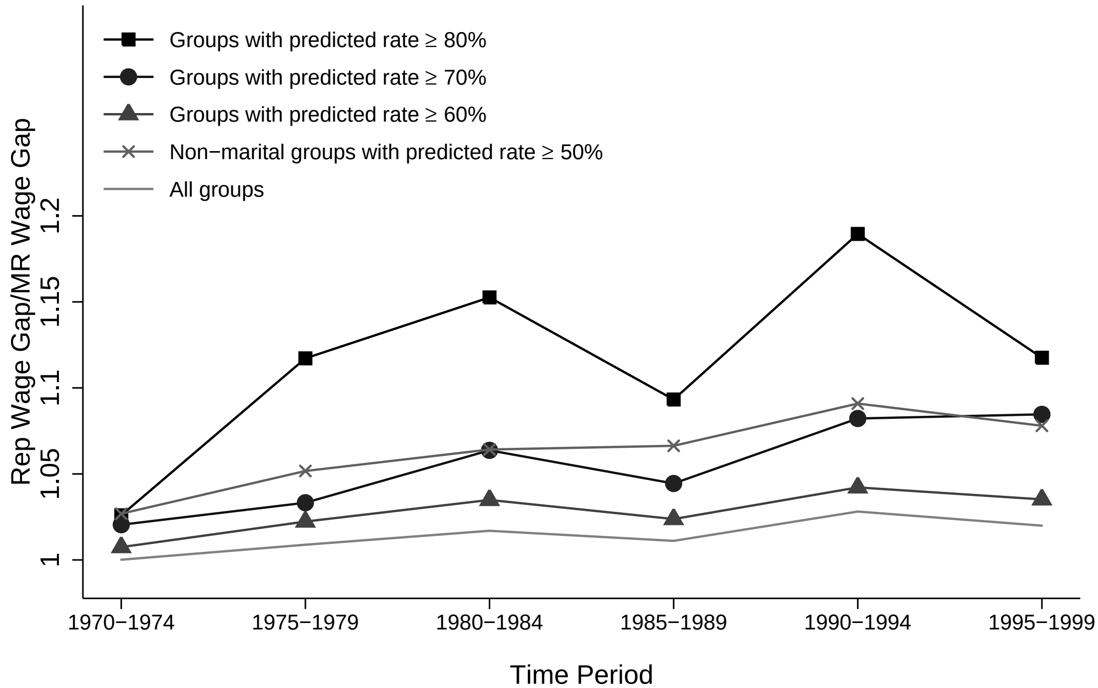


Figure 1.C.1: Relative gender wage gaps among strongly attached groups, various thresholds - Comparing the replication, using the CPS, to Mulligan and Rubinstein (2008)

The figure graphs five time series comparing the replication and original Mulligan and Rubinstein (2008) gender wage gaps, $\hat{\gamma}_t$, estimated in Section 1.3.3. The y-axis is the gender wage gap estimated in the replication divided by the gender wage gap estimated in Mulligan and Rubinstein (2008). The differences in the estimated gender wage gaps are caused by differences between the wage cutoffs used in the replication and in Mulligan and Rubinstein (2008).
The series differ according to the demographic groups (defined according to gender, schooling, marital status, and potential experience, except for the x-marked series which does not use marital status) included in the estimation. The unmarked series includes all demographic groups to calculate the wage ratio over time. For the other series, demographic groups are selected based on their full-time full-year labor force participation rate, which is predicted using the first-stage coefficients estimated with 1975 to 1979 (work-year) CPS data. The calculations use the CPS wage sample of white, non-Hispanic individuals aged 26 to 55.

Figure 1.C.2: Gender wage gap and wage inequality time series using the CPS

Using the 1971 to 2015 CPS ASEC samples, the figure compares how the median log hourly wage of women relative to the median log hourly wage of men has changed over time (left y-axis). It also displays a time series of the difference in the 90^{th} and 10^{th} percentiles of log hourly wages, a measure of wage inequality, by gender (right y-axis). The sample consists of white, non-Hispanic individuals aged 26 to 55, not living in group quarters.

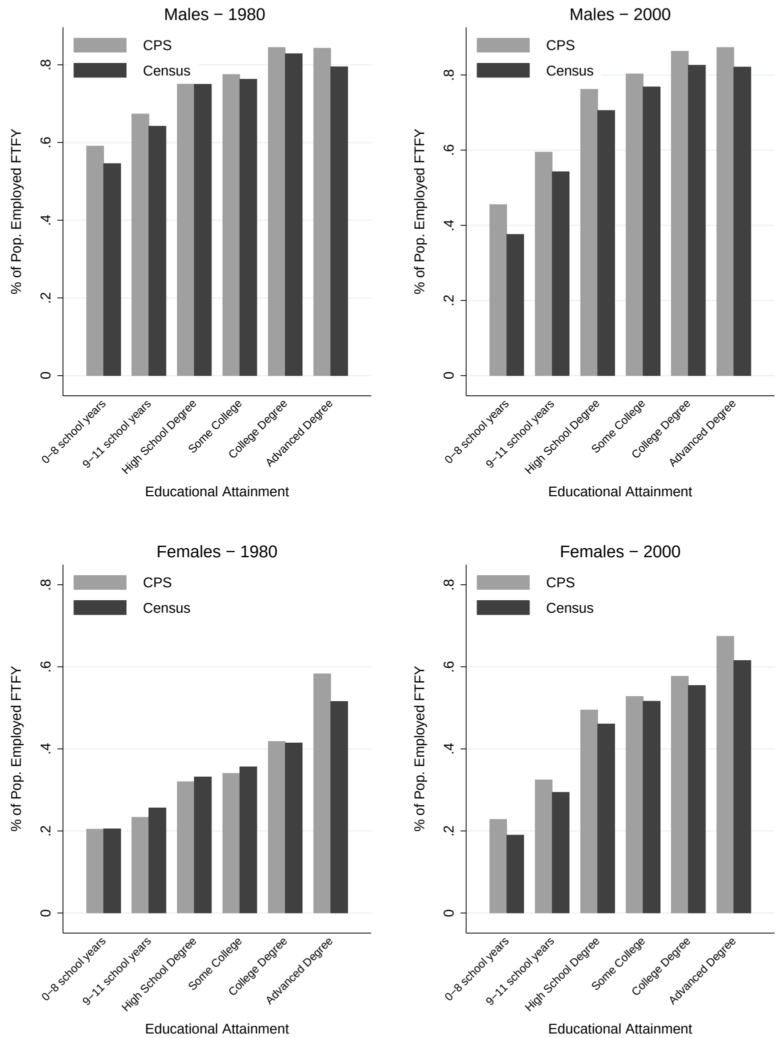


Figure 1.C.3: Percentage employed FTFY by educational attainment

The figure compares how the percentage of men and women employed full-time full-year (FTFY) by educational attainment differs between the CPS and the Census. Graphs labeled 1980 use the 1976 to 1980 (survey-year) CPS samples and the 1980 full Census sample. Graphs labeled 2000 use the 1996 to 2000 (survey-year) CPS samples and the 2000 full Census sample. The percentages are calculated using the CPS and Census samples of white, non-Hispanic individuals aged 26 to 55. I use a modified definition of full-time full-year labor force participation, as described in Section 1.4.2.2. The original definition of full-time full-year is working at least 50 weeks per year and at least 35 hours per week. The definition is then modified so that teachers or professors working at least 36 weeks per year and at least 35 hours per week are classified as full-time full-year. All other occupations use the original definition.

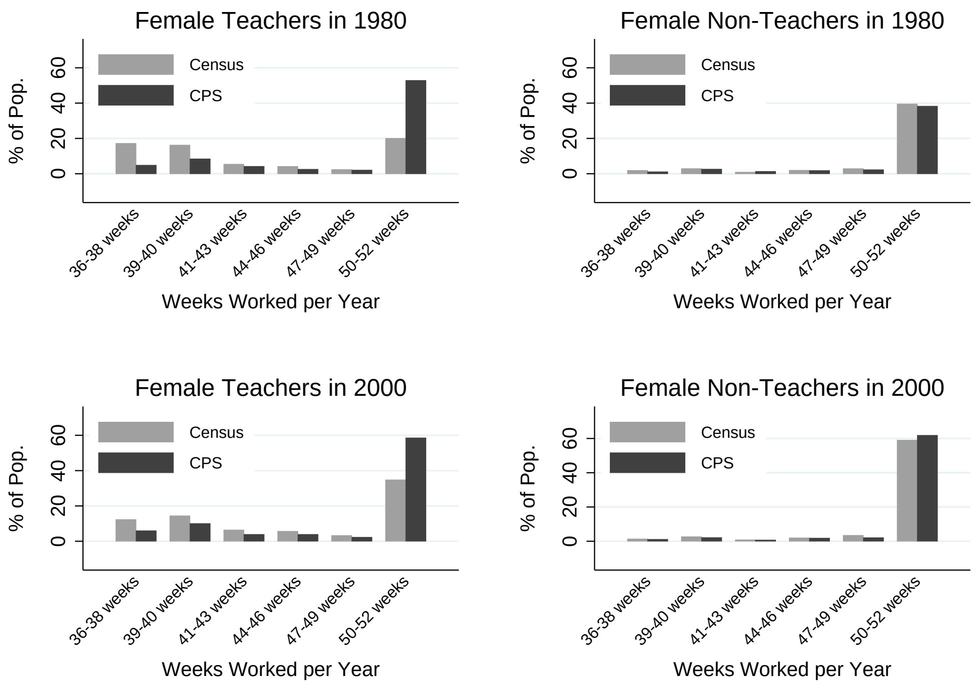


Figure 1.C.4: Weeks worked per year for female teachers and non-teachers in 1980 and 2000

These graphs compare how the distribution of weeks worked per year for employed females differs between the CPS and "CPS-Similar" samples. The graphs are separated by occupation, namely, teachers and non-teachers. Non-teachers include all occupations except teaching related positions. Graphs labeled 1980 use the 1976 to 1980 (survey-year) CPS samples and the 1980 "CPS-Similar" sample. Graphs labeled 2000 use the 1996 to 2000 (survey-year) CPS samples and the 2000 "CPS-Similar" sample. The CPS and Census samples consist of white, non-Hispanic women aged 26 to 55, not residing in group quarters.

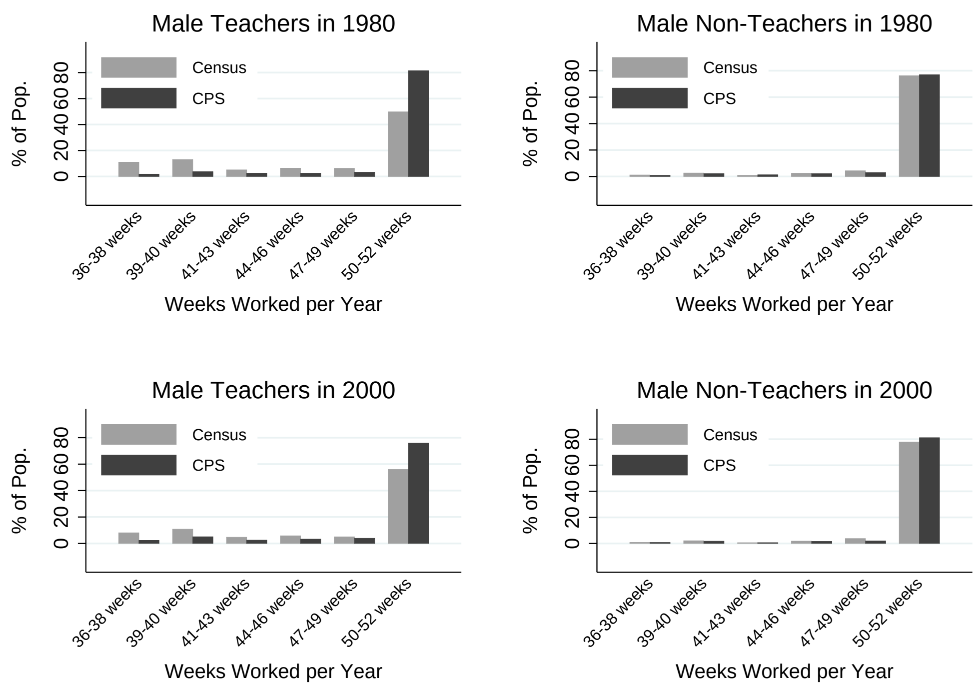


Figure 1.C.5: Weeks worked per year for male teachers and non-teachers in 1980 and 2000

These graphs compare how the distribution of weeks worked per year for employed males differs between the CPS and "CPS-Similar" samples. The graphs are separated by occupation, namely, teachers and non-teachers. Non-teachers include all occupations except teaching related positions. Graphs labeled 1980 use the 1976 to 1980 (survey-year) CPS samples and the 1980 "CPS-Similar" sample. Graphs labeled 2000 use the 1996 to 2000 (survey-year) CPS samples and the 2000 "CPS-Similar" sample. The CPS and Census samples consist of white, non-Hispanic men aged 26 to 55, not residing in group quarters.

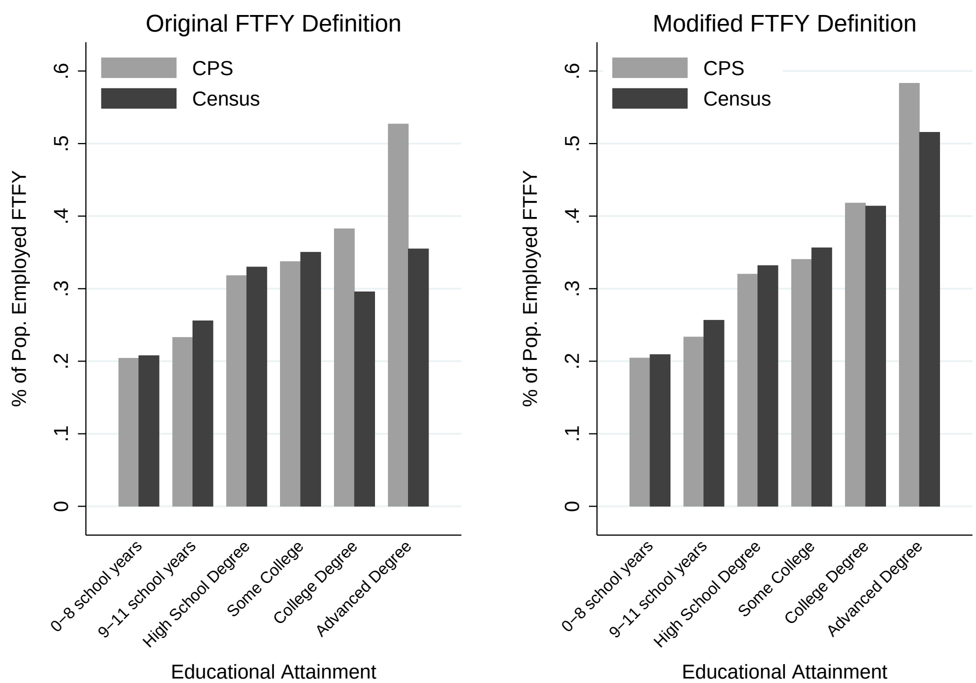


Figure 1.C.6: Percentage of females employed FTFY by educational attainment - Comparing the 1975-1979 CPS and 1980 "CPS-Similar" sample

These graphs show how changing the definition of full-time full-year makes the 1975-1979 CPS and 1980 "CPS-Similar" female labor force samples more comparable. The original definition of full-time full-year is working at least 50 weeks per year and at least 35 hours per week. The definition is then modified so that teachers or professors working at least 36 weeks per year and at least 35 hours per week are classified as full-time full-year. The modified definition is used in both the CPS and "CPS-Similar" samples. All other occupations use the original definition.

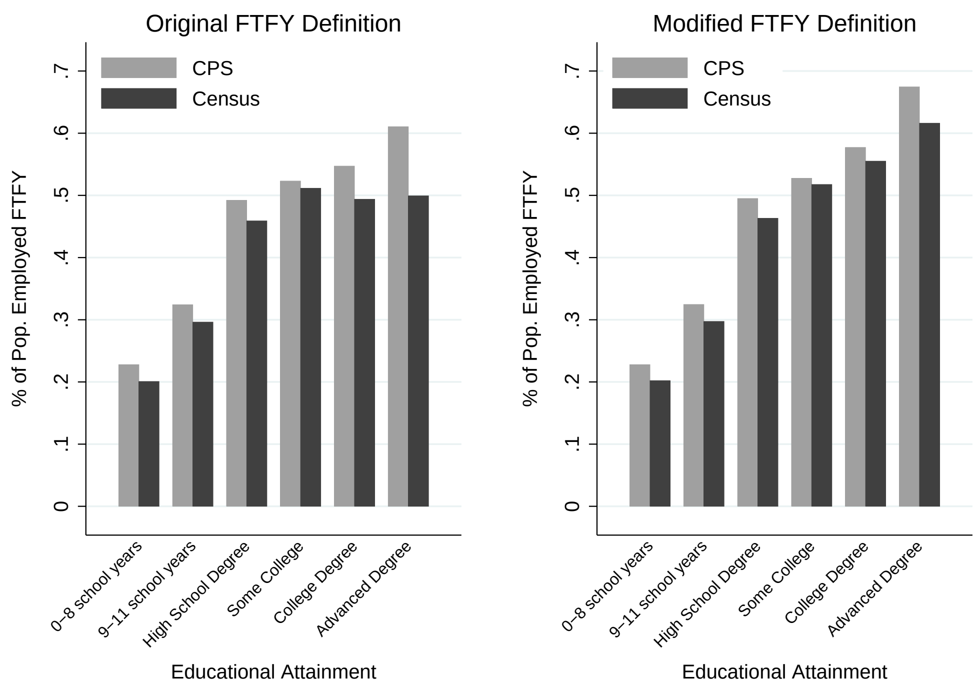


Figure 1.C.7: Percentage of females employed FTFY by educational attainment - Comparing the 1995-1999 CPS and 2000 "CPS-Similar" sample

These graphs show how changing the definition of full-time full-year makes the 1995-1999 CPS and 2000 "CPS-Similar" female labor force samples more comparable. The original definition of full-time full-year is working at least 50 weeks per year and at least 35 hours per week. The definition is then modified so that teachers or professors working at least 36 weeks per year and at least 35 hours per week are classified as full-time full-year. The modified definition is used in both the CPS and "CPS-Similar" samples. All other occupations use the original definition.

Table 1.C.1: Effects on the probability of working FTFY - Using the CPS to control for female selection, 1975-1979 and 1995-1999

	MR (2008)				Replication			
	1975-1979 Female (1)	1995-1999 Female (2)	1975-1979 Female (3)	1975-1979 Male (4)	1975-1979 Female (5)	1995-1999 Female (6)	1975-1979 Female (7)	1975-1979 Male (8)
Total Observations	116,843	102,395	116,843	112,291	116,843	102,395	116,843	112,291
Predicted *Pr*	0.297*	0.510*	0.307*	0.771*	0.298*	0.510*	0.308*	0.771*
Widowed	0.135*	0.024*	0.148*	-0.112*	0.136*	0.024	0.148*	-0.112*
Divorced	0.269*	0.156*	0.306*	-0.162*	0.269*	0.156*	0.306*	-0.162*
Separated	0.128*	0.065*	0.156*	-0.157*	0.128*	0.065*	0.156*	-0.157*
Never Married	0.284*	0.121*	0.375*	-0.181*	0.284*	0.121*	0.374*	-0.181*
Children 0-6 x constant	-0.183*	-0.158*			-0.183*	-0.158*		
widowed	-0.047	0.090*			-0.046	0.090		
divorced	0.008	0.032*			0.008	0.032*		
separated	0.021	0.011			0.021	0.011		
never-married	-0.062	-0.011			-0.063	-0.010		
Midwest	0.029*	0.036*	0.028*	0.019*	0.029*	0.036*	0.028*	0.019*
South	0.071*	0.037*	0.076*	-0.009*	0.071*	0.037*	0.076*	-0.010*
West	0.001	-0.037*	0.007	-0.058*	0.001	-0.037*	0.007	-0.058*
0-8 school years	-0.153*	-0.303*	-0.188	-0.295*	-0.153*	-0.303*	-0.188*	-0.295*
9-11 school years	-0.098*	-0.228*	-0.111	-0.202*	-0.098*	-0.228*	-0.111*	-0.202*
High School Degree	-0.010	-0.023*	-0.017	-0.061*	-0.010	-0.023*	-0.017*	-0.061*
College Degree	0.021*	0.021*	0.028	0.047*	0.020	0.021*	0.028*	0.047*
Advanced Degree	0.193*	0.109*	0.219	0.043*	0.193*	0.109*	0.218*	0.043*
$(exp-15)$	-0.005*	-0.005*	0.008	0.005*	-0.005*	-0.005*	0.008*	0.005*
$(exp-15)^2/100$	0.050*	0.058*	0.029	-0.064*	0.050*	0.058*	0.029*	-0.064*
$(exp-15)^3/1000$	-0.013	0.022	-0.076	0.031*	-0.012	0.022	-0.076*	0.031*
$(exp-15)^4/10000$	0.001	-0.024*	0.027	-0.010	0.001	-0.024*	0.027*	-0.010

The table lists the probit estimates of the probability of working full-time full-year (FTFY), evaluated at sample means. The first four columns are estimates taken from Mulligan and Rubinstein 2008. The last four columns list the replication estimates.
Following Mulligan and Rubinstein (2008), each column with estimates of the exclusion restriction, marital status interacted with the number of children under the age of 6, represents the first step of the Heckman two-step method to model female selection into the labor force. Fitted values from the columns that omit the exclusion restriction from estimation are used to select the samples for identification at infinity. All specifications include schooling-experience interactions (not shown). An asterisk (*) indicates statistically significant coefficients at the 95% confidence level. An asterisk beside the predicted probability of labor force participation indicates that the overall probit equation is statistically significant. The benchmark groups are currently married (living with spouse), some college, 15 years' potential experience, and Northeast region.

Table 1.C.2: Log wage regression coefficients for 1975-1979 - Using the CPS to control for female selection

	MR (2008)			Replication		
	1975-1979 Male (1)	1975-1979 Female (2)	1975-1979 Female (3)	1975-1979 Male (4)	1975-1979 Female (5)	1975-1979 Female (6)
Total Observations	57,457	27,656	27,656	58,003	27,578	27,578
Constant	3.027*	2.570*	2.665*	3.031*	2.572*	2.668*
Inverse Mills			-0.077*			-0.078*
Widowed	-0.030	0.012	-0.009	-0.036	0.009	-0.012
Divorced	-0.104*	0.041*	0.003	-0.107*	0.041*	0.002
Separated	-0.104*	-0.013	-0.033*	-0.112*	-0.014	-0.035*
Never Married	-0.181*	0.112*	0.068*	-0.180*	0.112*	0.068*
Midwest	0.008*	-0.056*	-0.060*	0.007	-0.056*	-0.060*
South	-0.088*	-0.112*	-0.121*	-0.087*	-0.111*	-0.121*
West	0.030*	0.002	0.002	0.030*	0.002	0.002
0-8 school years	-0.411*	-0.381*	-0.343*	-0.409*	-0.380*	-0.341*
9-11 school years	-0.272*	-0.292*	-0.273*	-0.273*	-0.293*	-0.273*
High School Degree	-0.091*	-0.109*	-0.106*	-0.091*	-0.106*	-0.103*
College Degree	0.190*	0.198*	0.195*	0.200*	0.198*	0.194*
Advanced Degree	0.192*	0.324*	0.299*	0.231*	0.336*	0.310*
$(exp-15)$	0.013*	-0.002	-0.002	0.013*	-0.002	-0.002
$(exp-15)^2/100$	-0.118*	-0.018	-0.023	-0.110*	-0.017	-0.023
$(exp-15)^3/1000$	0.025	0.072*	0.079*	0.030	0.068*	0.075*
$(exp-15)^4/10000$	0.006	-0.027*	-0.029*	0.001	-0.025*	-0.028*

The table presents coefficients from regressions on log hourly wages using the 1975 to 1979 (work-year) CPS data. The first three columns are estimates taken from Mulligan and Rubinstein (2008). The last three columns list the replication results. Columns without a coefficient on the inverse Mills' ratio are estimates from OLS regressions, while columns with a coefficient on the inverse Mills' ratio are estimates from the Heckman two-step regressions.

Following Mulligan and Rubinstein (2008), all specifications include schooling-experience interactions (not shown). An asterisk (*) indicates statistically significant coefficients at the 95% confidence level. The benchmark groups are currently married (living with spouse), some college, 15 years' potential experience, and Northeast region.

Table 1.C.3: Log wage regression coefficients for 1995-1999 - Using the CPS to control for female selection

	MR (2008)			Replication		
	1995-1999 Male (1)	1995-1999 Female (2)	1995-1999 Female (3)	1995-1999 Male (4)	1995-1999 Female (5)	1995-1999 Female (6)
Total Observations	54,565	41,062	41,062	55,015	41,004	41,004
Constant	2.929*	2.663*	2.559*	2.937*	2.670*	2.560*
Inverse Mills			0.116*			0.124*
Widowed	-0.138*	-0.079*	-0.072*	-0.140*	-0.079*	-0.071*
Divorced	-0.131*	-0.010	0.022*	-0.133*	-0.010	0.024*
Separated	-0.078*	-0.061*	-0.047*	-0.080*	-0.063*	-0.047*
Never Married	-0.186*	-0.003	0.027*	-0.189*	-0.004	0.028*
Midwest	-0.059*	-0.099*	-0.093*	-0.065*	-0.101*	-0.095*
South	-0.096*	-0.124*	-0.117*	-0.098*	-0.124*	-0.116*
West	-0.020*	-0.027*	-0.034*	-0.022*	-0.026*	-0.033*
0-8 school years	-0.437*	-0.501*	-0.566*	-0.401*	-0.493*	-0.562*
9-11 school years	-0.435*	-0.421*	-0.465*	-0.433*	-0.416*	-0.463*
High School Degree	-0.174*	-0.214*	-0.219*	-0.174*	-0.212*	-0.217*
College Degree	0.277*	0.297*	0.302*	0.290*	0.302*	0.307*
Advanced Degree	0.398*	0.476*	0.496*	0.437*	0.491*	0.511*
$(exp-15)$	0.017*	0.009*	0.009*	0.017*	0.008*	0.009*
$(exp-15)^2/100$	-0.091*	-0.063*	-0.052*	-0.092*	-0.067*	-0.054*
$(exp-15)^3/1000$	0.023	0.053*	0.048*	0.019	0.057*	0.052*
$(exp-15)^4/10000$	-0.003	-0.022	-0.023*	0.000	-0.023*	-0.024*

The table presents coefficients from regressions on log hourly wages using the 1995 to 1999 (work-year) CPS data. The first three columns are estimates taken from Mulligan and Rubinstein (2008). The last three columns list the replication results. Columns without a coefficient on the inverse Mills' ratio are estimates from OLS regressions, while columns with a coefficient on the inverse Mills' ratio are estimates from the Heckman two-step regressions.

Following Mulligan and Rubinstein (2008), all specifications include schooling-experience interactions (not shown). An asterisk (*) indicates statistically significant coefficients at the 95% confidence level. The benchmark groups are currently married (living with spouse), some college, 15 years' potential experience, and Northeast region.

Table 1.C.4: Effects on the probability of working FTFY - Using the "CPS-Similar" sample

	1980 Female (1)	2000 Female (2)	2010 Female (3)
Total Observations	1,705,976	2,162,273	416,743
Predicted *Pr*	0.313*	0.477*	0.501*
Widowed	0.115*	0.045*	0.010
Divorced	0.252*	0.166*	0.097*
Separated	0.127*	0.060*	-0.002
Never Married	0.227*	0.126*	0.065*
Children 0-6 x constant	-0.151*	-0.124*	-0.106*
Midwest	0.012*	0.029*	0.022*
South	0.053*	0.022*	0.016*
West	-0.002*	-0.042*	-0.044*
0-8 school years	-0.132*	-0.313*	-0.333*
9-11 school years	-0.072*	-0.213*	-0.281*
High School Degree	0.004*	-0.046*	-0.073*
College Degree	-0.109*	-0.038*	0.013*
Advanced Degree	-0.040*	-0.017*	0.094*
$(exp-15)$	0.005*	-0.013*	0.002

The table uses the 1980, 2000, and 2010 "CPS-Similar" samples (see Section 1.4.1.1). Listed are the Heckman two-step probit estimates of the probability of working full-time full-year (FTFY), evaluated at sample means. The exclusion restriction is marital status interacted with the number of children under the age of 6. Following Mulligan and Rubinstein (2008), all specifications include schooling-experience interactions (not shown). Wage cutoffs are generated using Census data, following the wage trimming process described in Mulligan and Rubinstein (2008). An asterisk (*) indicates statistically significant coefficients at the 95% confidence level. An asterisk beside the predicted probability of labor force participation indicates that the overall probit equation is statistically significant. The benchmark groups are currently married (living with spouse), some college, 15 years' potential experience, and Northeast region.

Table 1.C.5: Log wage regression coefficients - Using the "CPS-Similar" sample

	1980 Male (1)	1980 Female (2)	1980 Female (3)	2000 Male (4)	2000 Female (5)	2000 Female (6)	2010 Male (7)	2010 Female (8)	2010 Female (9)
Total Observations	879,403	433,145	433,145	1,011,829	734,504	734,504	199,219	164,622	164,622
Constant	3.038*	2.585*	2.664*	2.964*	2.693*	2.594*	2.943*	2.677*	2.567*
Inverse Mills			-0.066*			0.114*			0.126*
Widowed	-0.085*	-0.010*	-0.024*	-0.129*	-0.056*	-0.050*	-0.122*	-0.100*	-0.100*
Divorced	-0.095*	0.033*	0.001	-0.127*	-0.009*	0.021*	-0.137*	-0.042*	-0.022*
Separated	-0.072*	-0.009*	-0.026*	-0.121*	-0.079*	-0.068*	-0.121*	-0.121*	-0.120*
Never Married	-0.188*	0.090*	0.058*	-0.184*	0.005*	0.034*	-0.201*	-0.049*	-0.030*
Midwest	0.022*	-0.028*	-0.029*	-0.071*	-0.112*	-0.107*	-0.130*	-0.136*	-0.132*
South	-0.074*	-0.085*	-0.091*	-0.107*	-0.117*	-0.113*	-0.116*	-0.126*	-0.123*
West	0.033*	0.021*	0.022*	-0.014*	-0.015*	-0.022*	-0.014*	-0.008	-0.016*
0-8 school years	-0.426*	-0.372*	-0.342*	-0.466*	-0.527*	-0.605*	-0.451*	-0.607*	-0.700*
9-11 school years	-0.276*	-0.293*	-0.277*	-0.377*	-0.450*	-0.498*	-0.390*	-0.414*	-0.486*
High School Degree	-0.103*	-0.132*	-0.130*	-0.180*	-0.211*	-0.222*	-0.193*	-0.218*	-0.235*
College Degree	0.226*	0.206*	0.220*	0.323*	0.346*	0.343*	0.356*	0.381*	0.392*
Advanced Degree	0.275*	0.334*	0.339*	0.451*	0.505*	0.505*	0.545*	0.584*	0.607*
$(exp-15)$	0.188*	0.031*	0.031*	0.166*	0.115*	0.117*	0.177*	0.123*	0.133*
$(exp-15)^2/100$	-0.101*	-0.037*	-0.043*	-0.096*	-0.068*	-0.061*	-0.065*	-0.042*	-0.037*
$(exp-15)^3/1000$	-0.015*	0.030*	0.033*	0.007	0.026*	0.027*	0.013	0.022	0.017
$(exp-15)^4/10000$	0.014*	-0.008*	-0.008*	0.008*	-0.006*	-0.009*	-0.005	-0.007	-0.007

The table uses the 1980, 2000, and 2010 "CPS-Similar" samples (see Section 1.4.1.1). Listed are the coefficients from regressions on log hourly wages. Columns without a coefficient on the inverse Mills' ratio are estimates from OLS regressions, while columns with a coefficient on the inverse Mills' ratio are estimates from the Heckman two-step regressions.

Following Mulligan and Rubinstein (2008), all specifications include schooling-experience interactions (not shown). Wage cutoffs are generated using Census data, following the wage trimming process described in Mulligan and Rubinstein (2008). An asterisk (*) indicates statistically significant coefficients at the 95% confidence level. The benchmark groups are currently married (living with spouse), some college, 15 years' potential experience, and Northeast region.

Table 1.C.6: Gender-gap changes by marital status and schooling - Using the "CPS-Similar" sample to control for female selection, 1980 v 2010

	OLS			Two-Step	Bias
	1980 (1)	2010 (2)	Change (3)	Change (4)	Change (5)
			Panel A: All		
Conditional on marital status	-0.420	-0.236	0.184	0.008	0.176
	(0.001)	(0.004)	(0.004)	(0.015)	(0.016)
Not conditional on marital status	-0.446	-0.237	0.210	-0.040	0.250
	(0.001)	(0.004)	(0.004)	(0.012)	(0.013)
			Panel B: By Marital Status		
Currently Married	-0.487	-0.276	0.211	0.018	0.193
	(0.001)	(0.004)	(0.004)	(0.016)	(0.017)
Separated	-0.423	-0.275	0.148	-0.028	0.176
	(0.005)	(0.013)	(0.014)	(0.021)	(0.026)
Widowed	-0.411	-0.254	0.158	-0.022	0.180
	(0.008)	(0.023)	(0.025)	(0.027)	(0.037)
Divorced	-0.359	-0.181	0.178	0.036	0.142
	(0.002)	(0.007)	(0.008)	(0.015)	(0.017)
Never Married	-0.208	-0.123	0.085	-0.057	0.142
	(0.002)	(0.005)	(0.006)	(0.014)	(0.016)
			Panel C: By Education		
0 to 8 years	-0.363	-0.411	-0.048	-0.346	0.298
	(0.016)	(0.124)	(0.125)	(0.131)	(0.181)
High School, not grad	-0.452	-0.260	0.192	-0.057	0.249
	(0.005)	(0.028)	(0.029)	(0.031)	(0.042)
High school graduates	-0.443	-0.252	0.191	0.011	0.180
	(0.001)	(0.006)	(0.006)	(0.017)	(0.018)
Some College	-0.411	-0.226	0.186	0.023	0.163
	(0.002)	(0.003)	(0.004)	(0.014)	(0.015)
College	-0.439	-0.218	0.221	0.053	0.168
	(0.004)	(0.005)	(0.006)	(0.015)	(0.016)
Advanced Degree	-0.314	-0.198	0.116	-0.038	0.154
	(0.011)	(0.010)	(0.014)	(0.019)	(0.024)

Using the "CPS-Similar" sample, each cell summarizes the regression results listed in Tables 1.C.4 and 1.C.5. The entries are female minus male log wages, which differ from each other in terms of (a) rows, or demographic groups; (b) columns, or the time period used for estimation (1980 and 2010) and whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or the types of demographic groups summarized. Time-invariant, or fixed, weights are used to calculate log wages. The "Bias" column is the difference between the OLS and two-step columns. Bootstrap standard errors are in parentheses.

Table 1.C.7: Effects on the probability of working FTFY - Using the "CPS-Similar" and full Census sample to model selection

	CPS-Similar Sample			Full Sample		
	1980 Female (1)	1980 Male (2)	2010 Male (3)	1980 Female (4)	1980 Male (5)	2010 Male (6)
Total Observations	1,705,976	1,652,348	398,870	1,716,016	1,672,448	407,486
Predicted *Pr*	0.337*	0.734*	0.691*	0.337*	0.741*	0.679*
Widowed	0.113*	-0.131*	-0.173*	0.111*	-0.149*	-0.193*
Divorced	0.286*	-0.126*	-0.143*	0.283*	-0.145*	-0.158*
Separated	0.144*	-0.112*	-0.157*	0.142*	-0.127*	-0.173*
Never Married	0.317*	-0.177*	-0.207*	0.301*	-0.207*	-0.223*
Midwest	0.020*	-0.001	-0.015*	0.020*	-0.001	-0.015*
South	0.072*	-0.023*	-0.004	0.072*	-0.008*	-0.006*
West	-0.002	-0.077*	-0.058*	-0.002	-0.063*	-0.060*
0-8 school years	-0.188*	-0.267*	-0.309*	-0.194*	-0.297*	-0.332*
9-11 school years	-0.104*	-0.170*	-0.292*	-0.105*	-0.188*	-0.306*
High School Degree	-0.011*	-0.047*	-0.086*	-0.011*	-0.049*	-0.091*
College Degree	0.022*	0.047*	0.088*	0.022*	0.047*	0.094*
Advanced Degree	0.145*	0.014*	0.121*	0.143*	0.019*	0.129*
$(exp-15)$	0.075*	0.032*	-0.011*	0.073*	0.019*	-0.010*

The table lists the probit estimates of the probability of working full-time full-year (FTFY), evaluated at sample means. The modified definition of full-time full-year is used. Fitted values from the columns are used to select the samples for identification at infinity. Note, the exclusion restriction (the number of children under the age of 6 interacted with marital status) is omitted. The first three columns use the "CPS-Similar" sample. The last three columns use the full Census sample, which includes individuals residing in group quarters. Columns 1 and 4 use the 1980 Census to model female selection into the labor force, and Columns 2 and 5 use the 1980 Census to model male selection into the labor force. Columns 3 and 6 use the 2010 ACS to model male selection into the labor force. Wage cutoffs are generated using Census data, following the wage trimming process described in Mulligan and Rubinstein (2008). Following Mulligan and Rubinstein (2008), all specifications include schooling-experience interactions (not shown). An asterisk (*) indicates statistically significant coefficients at the 95% confidence level. An asterisk beside the predicted probability of labor force participation indicates that the overall probit equation is statistically significant. The benchmark groups are currently married (living with spouse), some college, 15 years' potential experience, and Northeast region.

Table 1.C.8: Marital status and educational attainment groups, 50% threshold - From the first-stage identification at infinity selection equation

	Men				Women			
	CPS	CPS-Similar,1980	Full Census,1980	Full Census, 2010	CPS	CPS-Similar,1980	Full Census,1980	Full Census, 2010
Panel A: Marital Status								
Currently Married	X	X	X	X	X	X	X	X
Separated	X	X	X	X	X	X	X	X
Widowed	X	X	X	X	X	X	X	X
Divorced	X	X	X	X	X	X	X	X
Never Married	X	X	X	X	X	X	X	X
Panel B: Educational Attainment								
0-8 school years	X	X	X					
9-11 school years	X	X	X					
High School Degree	X	X	X	X				
Some College	X	X	X	X	X (if year $\leq$ 1990)	X (if year $\leq$ 1980)	X (if year $\leq$ 1980)	X (if year $\leq$ 1980)
College Degree	X	X	X	X	X	X	X	X
Advanced Degree	X	X	X	X	X	X	X	X

The table lists the marital status and educational attainment groups, by sex, selected in the first stage of identification at infinity. Categories marked with an "X" indicate that to have a predicted full-time full-year labor force participation rate of at least 50 percent, an individual must be in the respective marital status or education group. However, the demographic groups for identification at infinity are not solely defined by one dimension, such as educational attainment or marital status. Rather they are defined by education x potential experience x region, such that $P(\mathbf{X}_{it}\hat{\delta}_g)$ is greater than α. Recall that Equation (1.3) uses all regressors other than marital status to estimate the gender wage gap when $\alpha = 50$.
Series marked with "X (year $\leq$ 1980)" indicate that there are individuals in that category who have predicted labor force participation rates above 50 percent in 1970 and 1980 only; there are no individuals in that category with strong predicted labor force attachment in later years. Similarly for series marked with "X (year $\leq$ 1990)."
The columns correspond to the following datasets: the CPS, the "CPS-Similar" sample, and the full Census sample. All samples use the 1980 Census to predict full-time full-year labor force participation for females. One instance uses the 2010 ACS to model male selection into the labor force; all others use the 1980 Census for men's first-stage estimates.

Table 1.C.9: Marital status and educational attainment groups, 60% threshold - From the first-stage identification at infinity selection equation

	Men				Women			
	CPS	CPS-Similar,1980	Full Census,1980	Full Census, 2010	CPS	CPS-Similar,1980	Full Census,1980	Full Census, 2010
Panel A: Marital Status								
Currently Married	X	X	X	X	X			
Separated	X	X	X	X	X	X	X	X
Widowed	X	X	X	X	X	X	X	X
Divorced	X	X	X	X	X	X	X	X
Never Married	X	X	X	X	X	X	X	X
Panel B: Educational Attainment								
0-8 school years	X	X	X					
9-11 school years	X	X	X		X			
High School Degree	X	X	X	X	X	X	X	X
Some College	X	X	X	X	X	X	X	X
College Degree	X	X	X	X	X	X	X	X
Advanced Degree	X	X	X	X	X	X	X	X

The table lists the marital status and educational attainment groups, by sex, selected in the first stage of identification at infinity. Categories marked with an "X" indicate that to have a predicted full-time full-year labor force participation rate of at least 60 percent, an individual must be in the respective marital status or education group. However, the demographic groups for identification at infinity are not solely defined by one dimension, such as educational attainment or marital status. Rather they are defined by marital status x education x potential experience x region, such that $P(\mathbf{X}_{it}\hat{\delta}_g)$ is greater than α.

The columns correspond to the following datasets: the CPS, the "CPS-Similar" sample, and the full Census sample. All samples use the 1980 Census to predict full-time full-year labor force participation for females. One instance uses the 2010 ACS to model male selection into the labor force; all others use the 1980 Census for men's first-stage estimates.

Table 1.C.10: Marital status and educational attainment groups, 70% threshold - From the first-stage identification at infinity selection equation

	Men				Women			
	CPS	CPS-Similar,1980	Full Census,1980	Full Census, 2010	CPS	CPS-Similar,1980	Full Census,1980	Full Census, 2010
Panel A: Marital Status								
Currently Married	X	X	X	X				
Separated	X	X	X	X	X			
Widowed	X	X	X	X	X			
Divorced	X	X	X	X	X	X	X	X
Never Married	X	X	X	X	X	X	X	X
Panel B: Educational Attainment								
0-8 school years								
9-11 school years	X	X	X					
High School Degree	X	X	X	X	X			
Some College	X	X	X	X	X	X (if year $\leq$ 1980)	X (if year $\leq$ 1980)	X (if year $\leq$ 1980)
College Degree	X	X	X	X	X	X	X	X
Advanced Degree	X	X	X	X	X	X	X	X

The table lists the marital status and educational attainment groups, by sex, selected in the first stage of identification at infinity. Categories marked with an "X" indicate that to have a predicted full-time full-year labor force participation rate of at least 70 percent, an individual must be in the respective marital status or education group. However, the demographic groups for identification at infinity are not solely defined by one dimension, such as educational attainment or marital status. Rather they are defined by marital status x education x potential experience x region, such that $P(\mathbf{X}_{it}\hat{\delta}_g)$ is greater than α.

Series marked with "X (year $\leq$ 1980)" indicate that there are individuals in that category who have predicted labor force participation rates above 70 percent in 1970 and 1980 only; there are no individuals in that category with strong predicted labor force attachment in later years.

The columns correspond to the following datasets: the CPS, the "CPS-Similar" sample, and the full Census sample. All samples use the 1980 Census to predict full-time full-year labor force participation for females. One instance uses the 2010 ACS to model male selection into the labor force; all others use the 1980 Census for men's first-stage estimates.

Table 1.C.11: Marital status and educational attainment groups, 80% threshold - From the first-stage identification at infinity selection equation

	Men				Women			
	CPS	CPS-Similar,1980	Full Census,1980	Full Census, 2010	CPS	CPS-Similar,1980	Full Census,1980	Full Census, 2010
Panel A: Marital Status								
Currently Married	X	X	X	X				
Separated	X	X	X	X				
Widowed	X							
Divorced	X			X	X			
Never Married	X				X	X	X	X
Panel B: Educational Attainment								
0-8 school years								
9-11 school years								
High School Degree	X	X	X					
Some College	X	X	X	X				
College Degree	X	X	X	X	X	X (if year $\leq$ 1980)		
Advanced Degree	X	X	X	X	X	X	X	X

The table lists the marital status and educational attainment groups, by sex, selected in the first stage of identification at infinity. Categories marked with an "X" indicate that to have a predicted full-time full-year labor force participation rate of at least 80 percent, an individual must be in the respective marital status or education group. However, the demographic groups for identification at infinity are not solely defined by one dimension, such as educational attainment or marital status. Rather they are defined by marital status x education x potential experience x region, such that $P(\mathbf{X}_{it}\hat{\delta}_g)$ is greater than α.

Series marked with "X (year $\leq$ 1980)" indicate that there are individuals in that category who have predicted labor force participation rates above 80 percent in 1970 and 1980 only; there are no individuals in that category with strong predicted labor force attachment in later years.

The columns correspond to the following datasets: the CPS, the "CPS-Similar" sample, and the full Census sample. All samples use the 1980 Census to predict full-time full-year labor force participation for females. One instance uses the 2010 ACS to model male selection into the labor force; all others use the 1980 Census for men's first-stage estimates.

Table 1.C.12: Effects on the probability of working FTFY - Using the CPS and factor variables, 1975-1979 and 1995-1999

	1975-1979 Female (1)	1995-1999 Female (2)	1975-1979 Female (3)	1975-1979 Male (4)
Total Observations	116,843	102,395	116,843	112,291
Predicted *Pr*	0.317*	0.511*	0.317*	0.759*
Widowed	0.108*	0.053*	0.141*	-0.100*
Divorced	0.262*	0.168*	0.302*	-0.149*
Separated	0.126*	0.069*	0.149*	-0.142*
Never Married	0.248*	0.119*	0.372*	-0.170*
Children 0-6 x constant	-0.179*	-0.153*		
Midwest	0.027*	0.036*	0.027*	0.017*
South	0.068*	0.037*	0.076*	-0.009*
West	0.001	-0.037*	0.007	-0.056*
0-8 school years	-0.146*	-0.316*	-0.198*	-0.251*
9-11 school years	-0.072*	-0.205*	-0.097*	-0.176*
High School Degree	-0.004	-0.026*	-0.014	-0.042*
College Degree	0.021	0.003	0.030*	0.037*
Advanced Degree	0.200*	0.087*	0.213*	0.029*
$(exp-15)$	0.006	-0.008	0.077*	0.031*

The table lists the probit estimates of the probability of working full-time full-year (FTFY), evaluated at sample means. The probit estimates are calculated using data from Mulligan and Rubinstein 2008, using factor variables to create interactions rather than using pre-generated interactions.

Following Mulligan and Rubinstein (2008), each column with estimates of the exclusion restriction, marital status interacted with the number of children under the age of 6, represents the first step of the Heckman two-step method to model female selection into the labor force. Fitted values from the columns without the exclusion restriction are used to select the samples for identification at infinity. All specifications include schooling-experience interactions (not shown). An asterisk (*) indicates statistically significant coefficients at the 95% confidence level. An asterisk beside the predicted probability of labor force participation indicates that the overall probit equation is statistically significant. The benchmark groups are currently married (living with spouse), some college, 15 years' potential experience, and Northeast region.

Table 1.C.13: Correcting the gender wage gap using the Heckman two-step estimator - Using the CPS and factor variables, 1975-1979 and 1995-1999

Period	Method		Bias
	OLS (1)	Two-Step (2)	(3)
Panel A: Variable Weights			
1975-1979	-0.417	-0.339	-0.079
	(0.007)	(0.015)	(0.016)
1995-1999	-0.265	-0.358	0.092
	(0.004)	(0.015)	(0.016)
Change	0.152	-0.019	0.171
	(0.008)	(0.021)	(0.023)
Panel B: Fixed Weights			
1975-1979	-0.409	-0.335	-0.075
	(0.005)	(0.014)	(0.015)
1995-1999	-0.274	-0.371	0.096
	(0.005)	(0.016)	(0.017)
Change	0.135	-0.036	0.171
	(0.008)	(0.021)	(0.022)

As in Mulligan and Rubinstein (2008), each cell summarizes the regression results listed in Tables 1.C.2, 1.C.3, and 1.C.12. The entries are female minus male log wages, which differ from each other in terms of (a) rows, or time period used for estimation (1975-1979 vs 1995-1999); (b) columns, or whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or whether variable or fixed weights are used to calculate log wages (see Appendix 1.A). The "Bias" column is the difference between the OLS and two-step columns. The "Change" row is the difference between the 1995-1999 and 1975-1979 rows.

The regressions control for demographics interacted with sex and use the CPS wage sample of white, non-Hispanic individuals aged 26 to 55. The wage trimming process is described in Section 1.3.1. Factor variables are used to generate interaction terms. Bootstrap standard errors are in parentheses.

Table 1.C.14: Gender-gap changes by marital status and schooling - Using the CPS and factor variables, 1975-1979 and 1995-1999

	OLS			Two-Step	Bias
	1975-1979	1995-1999	Change	Change	Change
	(1)	(2)	(3)	(4)	(5)
			Panel A: All		
Conditional on marital status	-0.409	-0.274	0.135	-0.036	0.171
	(0.005)	(0.005)	(0.008)	(0.021)	(0.022)
Not conditional on marital status	-0.436	-0.281	0.155	-0.016	0.171
	(0.005)	(0.005)	(0.007)	(0.015)	(0.017)
			Panel B: By Marital Status		
Currently Married	-0.479	-0.323	0.156	-0.040	0.196
	(0.005)	(0.006)	(0.008)	(0.024)	(0.025)
Separated	-0.381	-0.306	0.075	-0.085	0.160
	(0.021)	(0.022)	(0.031)	(0.036)	(0.047)
Widowed	-0.434	-0.262	0.172	0.005	0.167
	(0.025)	(0.042)	(0.049)	(0.053)	(0.072)
Divorced	-0.331	-0.200	0.131	0.008	0.123
	(0.011)	(0.009)	(0.015)	(0.020)	(0.025)
Never Married	-0.187	-0.137	0.050	-0.070	0.119
	(0.010)	(0.010)	(0.014)	(0.020)	(0.024)
			Panel C: By Education		
0 to 8 years	-0.369	-0.271	0.098	-0.209	0.307
	(0.060)	(0.086)	(0.105)	(0.110)	(0.152)
High School, not grad	-0.442	-0.346	0.096	-0.142	0.237
	(0.020)	(0.037)	(0.042)	(0.050)	(0.065)
High school graduates	-0.429	-0.302	0.127	-0.046	0.173
	(0.006)	(0.008)	(0.010)	(0.022)	(0.024)
Some College	-0.412	-0.261	0.151	-0.014	0.165
	(0.007)	(0.007)	(0.010)	(0.022)	(0.024)
College	-0.432	-0.255	0.177	0.014	0.162
	(0.013)	(0.012)	(0.018)	(0.026)	(0.031)
Advanced Degree	-0.254	-0.198	0.056	-0.075	0.131
	(0.036)	(0.023)	(0.042)	(0.045)	(0.062)

As in Mulligan and Rubinstein (2008), each cell summarizes the regression results listed in Tables 1.C.2, 1.C.3, and 1.C.12. The entries are female minus male log wages, which differ from each other in terms of (a) rows, or demographic groups; (b) columns, or the time period used for estimation (1975-1979 and 1995-1999) and whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or the types of demographic groups summarized. Time-invariant, or fixed, weights are used to calculate log wages. The "Bias" column is the difference between the OLS and two-step columns.

The regressions control for demographics (which include marital status unless indicated otherwise) interacted with sex and use the CPS wage sample of white, non-Hispanic individuals aged 26 to 55. The wage trimming process is described in Section 1.3.1. Factor variables are used to generate interaction terms. Bootstrap standard errors are in parentheses.

Table 1.C.15: Descriptive statistics for the 2005-2009 and 2010-2014 CPS samples

	Full Sample				FTFY Sample			
	2005-2009 Female (1)	2010-2014 Female (2)	2005-2009 Male (3)	2010-2014 Male (4)	2005-2009 Female (5)	2010-2014 Female (6)	2005-2009 Male (7)	2010-2014 Male (8)
Total Observations	145,651	129,959	137,097	122,758	67,733	60,471	87,069	76,878
FTFY Indicator	0.525	0.514	0.755	0.726	1	1	1	1
Log Real Hourly Wage					2.694	2.702	2.913	2.901
Inverse Mills					0.720	0.738	0	0
Widowed	0.018	0.018	0.006	0.006	0.015	0.014	0.004	0.004
Divorced	0.138	0.132	0.115	0.113	0.160	0.148	0.102	0.101
Separated	0.024	0.024	0.019	0.020	0.023	0.022	0.018	0.017
Never Married	0.141	0.172	0.215	0.242	0.167	0.192	0.185	0.207
0-8 school years	0.010	0.009	0.013	0.012	0.003	0.002	0.006	0.005
9-11 school years	0.022	0.019	0.029	0.025	0.011	0.008	0.020	0.015
High School Degree	0.279	0.247	0.330	0.307	0.261	0.219	0.313	0.282
College Degree	0.249	0.267	0.228	0.242	0.267	0.299	0.251	0.276
Advanced Degree	0.110	0.128	0.097	0.104	0.130	0.155	0.105	0.116
Potential Experience - 15	5.342	5.056	5.462	5.189	5.115	4.718	4.909	4.680
Midwest	0.267	0.263	0.266	0.264	0.277	0.275	0.272	0.270
South	0.341	0.349	0.337	0.344	0.350	0.355	0.338	0.350
West	0.197	0.196	0.203	0.200	0.179	0.175	0.192	0.186
Kids 0-6	0.349	0.348	0.336	0.330	0.239	0.258	0.363	0.358

Using the 2006 to 2010 and 2011 to 2015 (survey-year) CPS ASEC samples, the sample consists of white, non-Hispanic individuals between the ages of 26 and 55. Following Mulligan and Rubinstein (2008), omitted groups are currently married, some college, 15 years' potential experience, and Northeast region. The first four columns list the descriptive statistics for the entire sample, and the last four columns list the descriptive statistics for the full-time full-year sample. Wage cutoffs are generated following the wage trimming process described in Mulligan and Rubinstein (2008).

Table 1.C.16: Effects on the probability of working FTFY and log wage regression coefficients - Using the CPS and factor variables, 2005-2009 and 2010-2014

	Marginal Effects		Log Wage Regressions					
	2005-2009 Female (1)	2010-2014 Female (2)	2005-2009 Male (3)	2005-2009 Female (4)	2005-2009 Female (5)	2010-2014 Male (6)	2010-2014 Female (7)	2010-2014 Female (8)
Total Observations	142,447	126,731	84,705	66,496	66,496	74,765	59,087	59,087
Predicted *Pr*/Constant	0.527*	0.515*	2.957*	2.676*	2.588*	2.908*	2.640*	2.497*
Inverse Mills					0.103*			0.159*
Widowed	-0.032*	-0.041*	-0.141*	-0.082*	-0.088*	-0.164*	-0.061*	-0.073*
Divorced	0.112*	0.093*	-0.143*	-0.026*	-0.008	-0.148*	-0.023*	0.001
Separated	0.032*	0.013	-0.159*	-0.081*	-0.075*	-0.150*	-0.063*	-0.058*
Never Married	0.081*	0.057*	-0.223*	-0.031*	-0.012	-0.191*	-0.029*	-0.007
Children 0-6 x constant	-0.128*	-0.108*						
Midwest	0.034*	0.036*	-0.087*	-0.112*	-0.107*	-0.093*	-0.101*	-0.093*
South	0.032*	0.028*	-0.078*	-0.102*	-0.097*	-0.086*	-0.097*	-0.090*
West	-0.031*	-0.038*	0.013*	0.008	0.004	-0.003	-0.010	-0.019*
0-8 school years	-0.273*	-0.357*	-0.478*	-0.403*	-0.445*	-0.449*	-0.421*	-0.561*
9-11 school years	-0.265*	-0.256*	-0.398*	-0.534*	-0.585*	-0.381*	-0.391*	-0.490*
High School Degree	-0.047*	-0.034*	-0.181*	-0.203*	-0.214*	-0.163*	-0.214*	-0.231*
College Degree	0.013	0.056*	0.305*	0.321*	0.326*	0.326*	0.344*	0.365*
Advanced Degree	0.088*	0.114*	0.463*	0.501*	0.515*	0.489*	0.505*	0.539*
$(exp-15)$	0.004	0.009	0.156*	0.142*	0.151*	0.123*	0.086*	0.098*
$(exp-15)^2/100$			-0.080*	-0.060*	-0.058*	-0.062*	-0.067*	-0.062*
$(exp-15)^3/1000$			-0.011	-0.015	-0.020	0.057*	0.055*	0.052*
$(exp-15)^4/10000$			0.010	0.009	0.010	-0.022	-0.016	-0.016

The table uses the 2006 to 2010 and 2011 to 2015 (survey-year) CPS ASEC samples. The first two columns list the Heckman two-step probit estimates of the probability of working full-time full-year (FTFY), evaluated at sample means. The exclusion restriction is marital status interacted with the number of children under the age of 6. Columns 3 through 8 present the coefficients from regressions on log hourly wages.

Following Mulligan and Rubinstein (2008), all specifications include schooling-experience interactions (not shown). An asterisk (*) indicates statistically significant coefficients at the 95% confidence level. An asterisk beside the predicted probability of labor force participation indicates that the overall probit equation is statistically significant. The benchmark groups are currently married (living with spouse), some college, 15 years' potential experience, and Northeast region. Factor variables are used to generate interaction terms.

Table 1.C.17: Correcting the gender wage gap using the Heckman two-step estimator - Using the CPS and factor variables, 1975-1979, 2005-2009, and 2010-2014

Period	Method		Bias
	OLS (1)	Two-Step (2)	(3)
Panel A: Variable Weights			
1975-1979	-0.417	-0.339	-0.079
	(0.007)	(0.016)	(0.017)
1995-1999	-0.265	-0.358	0.092
	(0.004)	(0.015)	(0.016)
2005-2009	-0.241	-0.315	0.073
	(0.004)	(0.014)	(0.014)
2010-2014	-0.240	-0.367	0.126
	(0.005)	(0.020)	(0.021)
Change	0.177	-0.028	0.205
	(0.009)	(0.026)	(0.027)
Panel B: Fixed Weights			
1975-1979	-0.409	-0.335	-0.075
	(0.005)	(0.014)	(0.015)
1995-1999	-0.274	-0.371	0.096
	(0.005)	(0.016)	(0.017)
2005-2009	-0.259	-0.337	0.079
	(0.006)	(0.015)	(0.016)
2010-2014	-0.270	-0.410	0.139
	(0.009)	(0.024)	(0.026)
Change	0.139	-0.075	0.214
	(0.011)	(0.028)	(0.030)

As in Mulligan and Rubinstein (2008), each cell summarizes the regression results listed in Tables 1.C.2, 1.C.3, and 1.C.12. The entries are female minus male log wages, which differ from each other in terms of (a) rows, or time period used for estimation (1975-1979, 1995-1999, 2005-2009, and 2010-2014); (b) columns, or whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or whether variable or fixed weights are used to calculate log wages (see Appendix 1.A). The "Bias" column is the difference between the OLS and two-step columns. The "Change" row is the difference between the 2010-2014 and 1975-1979 rows. Factor variables are used to generate interaction terms. Bootstrap standard errors are in parentheses.

Table 1.C.18: Gender-gap changes by marital status and schooling - Using the CPS and factor variables, 2005-2009 and 2010-2014

	2005-2009 CPS			2010-2014 CPS		
	OLS	Two-Step	Bias	OLS	Two-Step	Bias
	Change (1)	Change (2)	Change (3)	Change (4)	Change (5)	Change (6)
Panel A: All						
Conditional on marital status	0.151	-0.003	0.153	0.139	-0.075	0.214
	(0.008)	(0.021)	(0.022)	(0.011)	(0.028)	(0.030)
Not conditional on marital status	0.179	-0.029	0.208	0.168	-0.099	0.267
	(0.008)	(0.016)	(0.018)	(0.011)	(0.023)	(0.025)
Panel B: By Marital Status						
Currently Married	0.169	-0.003	0.172	0.162	-0.073	0.235
	(0.008)	(0.023)	(0.025)	(0.011)	(0.030)	(0.032)
Separated	0.163	0.014	0.149	0.149	-0.063	0.212
	(0.031)	(0.036)	(0.048)	(0.037)	(0.044)	(0.058)
Widowed	0.195	0.037	0.158	0.202	-0.026	0.228
	(0.043)	(0.047)	(0.064)	(0.053)	(0.060)	(0.080)
Divorced	0.141	0.025	0.116	0.134	-0.037	0.171
	(0.015)	(0.021)	(0.026)	(0.018)	(0.028)	(0.033)
Never Married	0.068	-0.042	0.111	0.029	-0.140	0.169
	(0.014)	(0.020)	(0.025)	(0.016)	(0.027)	(0.031)
Panel C: By Education						
0 to 8 years	-0.266	-0.486	0.220	-0.586	-0.879	0.293
	(0.196)	(0.198)	(0.279)	(0.358)	(0.367)	(0.512)
High School, not grad	0.122	-0.097	0.219	0.032	-0.298	0.330
	(0.041)	(0.050)	(0.065)	(0.076)	(0.088)	(0.117)
High school graduates	0.150	-0.008	0.158	0.154	-0.071	0.225
	(0.009)	(0.022)	(0.024)	(0.011)	(0.029)	(0.031)
Some College	0.185	0.035	0.151	0.187	-0.022	0.209
	(0.010)	(0.021)	(0.023)	(0.010)	(0.027)	(0.029)
College	0.192	0.050	0.142	0.193	0.006	0.186
	(0.016)	(0.024)	(0.029)	(0.017)	(0.028)	(0.032)
Advanced Degree	0.065	-0.050	0.115	0.030	-0.130	0.160
	(0.041)	(0.044)	(0.060)	(0.042)	(0.046)	(0.062)

Each cell summarizes the regression results listed in Tables 1.C.2, 1.C.3, and 1.C.12. The entries are changes to estimated gender wage gaps, which differ from each other in terms of (a) rows, or demographic groups; (b) columns, or the time period used for estimation and whether the regression includes the inverse Mills' ratio (OLS does not include it, two-step does); and (c) panels, or the types of demographic groups summarized. Columns 1 and 2 compare the 1975-1979 and 2005-2009 CPS samples. Columns 4 and 5 compare the 1975-1979 and 2010-2014 CPS samples. Time-invariant, or fixed, weights are used to calculate log wages. The "Bias" column is the difference between the OLS and two-step columns. Factor variables are used to generate interaction terms. Bootstrap standard errors are in parentheses.

Chapter 2

The Effects of Alabama's Presumptive Sentencing Guidelines

2.1 Introduction

In October of 2006, Alabama began using voluntary sentencing guidelines for a select group of property, drug, burglary, and personal offenses. The sentencing guidelines function through the completion of sentencing worksheets, which use a point based system to determine whether a convicted offender should serve a prison sentence and, if so, the guidelines suggest a sentence length range based on the offender's total score. Because worksheet usage varied across the state, in October of 2013, the sentencing guidelines for drug and property offenses became presumptive, or mandatory except for extenuating circumstances.

Alabama's sentencing guidelines were created to reduce the number of non-violent offenders' sentenced to prison and reduce sentence length for those that receive a prison sentence. Therefore, this paper uses state prison admissions data to study how the shift from voluntary to presumptive sentencing guidelines affected sentence length. Figure 2.1 presents a time-series of average sentence length among offenders admitted into prison for the commission of affected property and drug offenses. The figure presents results separately by race: the gray solid represents average sentence length among black offenders, and the black solid line represents average sentence length among white offenders. Quarter 0 denotes the quarter before the presumptive sentencing guidelines were implemented (July to September 2013). In Figure 2.1a, two years prior to the implementation of the presumptive sentencing guidelines, the average sentence length among property offenders was approximately 61 months. In the two years following the transition, the average sentence fell by 21 percent, to 48 months. Among white property offenders sentence length fell by 14 months; among black offenders sentence length fell by 11 months. In Figure 2.1b, average sentence length fell by 10 months among white drug offenders and by 12 months among black drug offenders.

The transition to presumptive guidelines in Alabama provides a setting where a difference-in-differences approach can be used by comparing the sentence length of presumptive worksheet offenses to the sentence length of those that remained voluntary. Yet, the guidelines aim to limit judicial discretion, which may change prosecutors' behavior; and they intend to reduce the use of incarceration among non-violent offenders. Because of data limitations, I estimate how, conditional on incarceration, the shift to presumptive sentencing guidelines may have affected offense and offender composition. A change in the use of incarceration would likely affect the type of offenders that receive a prison sentence, leading to changes in inmate or offense composition and confounding the effect on average sentence length.

I find that the changes in composition are not consistent across offense type and race. Following the shift to presumptive sentencing guidelines, there is no statistically significant difference in the average number of property offenders admitted into prison, but all else equal, the average number of white drug offenders declines. Furthermore, among white drug offenders, there is a statistically significant decline in the average severity of admitted offenses. And while I find statistically significant increases in the average and median age at admission following the transition to presumptive guidelines, there are no statistically significant changes in racial composition.

Next, I estimate how the sentencing guidelines affected average total sentence length, assessing whether effects differed across black and white offenders. I present estimates from a difference regression and a difference-in-differences regression. Using personal offenses covered by the voluntary guidelines as a control group, I find that after the guidelines became presumptive, average sentence length for property offenders fell 23 percent, as the guidelines had intended. Accounting for a difference in time trends across race, the transition does not have a statistically significant effect on the change in sentence length across white and black offenders. Relative to white personal offenders, sentence length among white drug offenders falls 11 percent. But, following the shift to presumptive sentencing guidelines, there is no statistically significant difference in the percent change in sentence length between black personal and drug offenders.

Lastly, I study how the effects of the transition varied across the distribution of sentence length. Figure 2.2 graphs a time-series of percentiles of sentence length by worksheet type. Figures 2.2a and 2.2b illustrate that the largest changes in sentence length occur in the 75^{th} and 90^{th} percentiles among both property and drug offenders, respectively. Note, that similar to average sentence length in Figure 2.1, declines in sentence length begin several quarters before the guidelines were adopted. Using quantile regressions, I find that among property and drug worksheet offenders, the shift to presumptive sentencing guidelines is associated with a decrease in sentence length at all percentiles except the 10^{th}. And as expected, some of the largest changes in total sentence length occur in the upper tail of the sentence length distribution. Furthermore, among drug offenders, all else equal, the percent decline in black offenders' sentence length at the 90^{th} percentile was larger relative to white offenders' change in sentence length.

The paper proceeds as follows: Section 2.2 describes the literature on state sentencing guidelines, and Section 2.3 provides background on the Alabama sentencing guidelines. Section 2.4 describes the prison admissions data I use to estimate the effects of the presumptive guidelines on sentence length. Section 2.5 analyzes how the shift to presumptive guidelines may have affected offense and offender composition. Section 2.6 studies the effects on average sentence length, and Section 2.7 uses quantile regression to study effects on sentence length across the distribution of sentence length. Section 2.8 concludes.

2.2 Literature Review

As of 2017, 17 states use sentencing guidelines, which fall along a continuum of voluntary to presumptive (Mitchell 2017). Voluntary guidelines serve as a suggestion for courts; judicial compliance is discretionary. Presumptive guidelines prescribe sentence length ranges that judges are required to adhere to, but they can make departures as necessary. Guidelines seek to limit judicial discretion and achieve uniform sentencing across similarly situated offenders. They aim to limit variability in sentencing decisions by guiding decisions based on legally relevant factors, like offense severity and prior offending, rather than extralegal factors, such as race or sex (Engen and Gainey 2000; Miethe 1987). Some states, like Alabama, also use the guidelines to reduce and predict prison populations by encouraging the use of alternative sanctions or recommending shorter sentence lengths for non-violent offenders. On the other hand, some states, like Pennsylvania, adopt guidelines because judges are viewed as too lenient (Marvell 1994).

In general, researchers find that consistent with the purpose of sentencing guidelines, offense severity and prior offending are the principal determinants of both sentence type and sentence length. Race and other extralegal factors also directly affect sentencing outcomes, although to a smaller degree (Dixon 1995; Miethe and Moore 1985; Moore and Miethe 1986). Instead, race has a relatively strong indirect effect, through criminal history, on the likelihood and duration of a prison sentence (Miethe and Moore 1985).

Furthermore, variation in sentence length and racial disparity is largest in sentencing decisions that involve the most discretion (Mitchell 2005; Pfaff 2006). Therefore, in Alabama, as sentencing gudielines became increasingly binding, the shift from voluntary to presumptive guidelines is expected to have a stronger impact on sentencing decisions and on any racially disparate sentences than when Alabama transitioned from no guidelines to voluntary sentencing guidelines.

Like this paper, Edwards, Rushin, and Colquitt (2019) also studies the shift in Alabama's sentencing guidelines. Using trial data, the authors combine the incarceration and sentence length decisions into one measure and analyze how this sentencing outcome is affected by both the shift from no guidelines to voluntary sentencing guidelines and the shift from voluntary to presumptive guidelines.[1] Using a difference-in-differences model, the authors find that the transition from voluntary to presumptive guidelines had a stronger impact on sentencing decisions than the shift from no guidelines to voluntary guidelines. They find that when the sentencing guidelines became presumptive for non-violent property and drug offenses, average sentencing fell between 31 and 40 percent (or by 24 to 32 months). They also find that prior to the adoption of any guidelines, average sentencing was larger for non-violent black offenders, relative to non-violent white offenders. The voluntary guidelines did not affect racial disparity in sentencing, but the shift to presumptive sentencing guidelines closed the racial gap in sentencing by about 8 months.

Because the Alabama sentencing guidelines successfully altered judicial behavior, Edwards, Rushin, and Colquitt (2019) studies how sentencing decisions changed across judges, who are divided into quartiles based on sentencing toughness. Following the shift from no guidelines to voluntary sentencing guidelines, the effect on sentencing decisions was largest among judges in the middle quartiles of sentencing toughness. After the transition to presumptive sentencing guidelines, the toughest judges decreased sentence length the most.

Since the difference in average sentencing between the most lenient and toughest judges fell, the shift to presumptive guidelines may have had larger effects in the right tail of the sentence length distribution. Additionally, Britt (2009) describes how a differential effect of offense severity and prior record is built into sentencing guidelines. He finds that in the left tail of the

[1] If an individual receives probation (no confinement), they are included in the sample as a zero. Over 26 percent of their sample receives probation or community corrections.

sentence length distribution, where sentences are relatively short, the effects of the independent variables are typically small because the range of possible sentence lengths is small for less serious offenses and offenders. At higher levels of the sentence length distribution, the effects of legal and extralegal factors increase in magnitude because as the range of possible sentence lengths increases, these factors can play a larger role in sentence length decisions. Therefore, in addition to analyzing changes in average sentence length, I study how the transition from voluntary to presumptive sentencing guidelines may have had different effects across the sentence length distribution.

This paper contributes to the literature on sentencing guidelines by providing an expansive assessment on how the transition from voluntary to presumptive sentencing guidelines affected Alabama's prison population. While Edwards, Rushin, and Colquitt (2019) studies a measure of sentencing that combines the sentence disposition and sentence length decisions, this paper analyzes how the shift to presumptive guidelines may have affected sentence length (conditional on incarceration). Relative to the results found in Edwards, Rushin, and Colquitt (2019), my difference-in-differences estimates are smaller in magnitude (a decline between 30 and 41 percent relative to a decline between 11 and 23 percent, respectively). This suggests that in addition to a decline in sentence length, there may have been a reduction in the number of offenders receiving a prison sentence.

This paper also studies changes in total sentence length across race. While I do not find any statistically significant differences in the percent decline in average sentence length across black and white offenders, there is a statistically significant difference among drug worksheet offenders. With data on prior offending, researchers could better assess the role race plays in the distribution of sentence length decisions as judicial discretion becomes increasingly limited in Alabama.

2.3 Alabama Sentencing Guidelines

In 2005, Alabama's prisons were 213 percent over design capacity, and incarceration alternatives, such as community supervision and drug and alcohol treatment programs, were often underutilized (Alabama Sentencing Commission 2006). Therefore, in October of 2006, the Alabama Sentencing Commission (ASC) implemented the Initial Voluntary Sentencing Standards (Act No. 2006-312). The goals of the sentencing guidelines are: (1) reduce prison overcrowding by incarcerating fewer non-violent offenders and by reducing sentence lengths for non-violent offenders who receive a prison sentence, (2) improve the predictability of correctional populations, and (3) eliminate unwarranted sentencing disparity among individuals convicted of the same crime and with similar criminal histories.

Adherence to the sentencing standards occurs through the completion of a set of worksheets, which are designed to imitate the two primary decisions in criminal sentencing. The first worksheet, the Prison In/Out worksheet, suggests a sentence disposition, or whether the sentence is served in prison or through alternative sanctions. The second worksheet, the Prison Sentence Length worksheet (and associated Prison Sentence Length Range table), recommends a sentence length range for the duration of the sentence.

The sentencing standards provide recommended sentence length ranges for 26 felony offenses, which account for 87 percent of the felony convictions and sentences imposed in Alabama between October 1998 and May 2003. These offenses are split into three categories: property (including burglary), drug, and personal. There is a Prison In/Out worksheet and Prison Sentence Length worksheet for each offense category. While the guidelines were created to reflect Alabama's historical sentencing practices, the sentence length ranges for the offenses covered by the worksheets are narrower than those under existing statutory law. The guidelines were also designed so that drug and property offenders are less likely to receive a prison sentence, while offenders convicted of personal offenses are more likely to receive a prison sentence.

To develop the worksheets, the ASC analyzed 14,000 pre- and post-sentence investigative reports to assess which offender and offense characteristics were statistically relevant for sentencing decisions (Alabama Sentencing Commission 2005). The worksheets include the following characteristics, or sentencing factors: the most serious current offense[2], additional offenses being sentenced at the current sentencing event, an individual's criminal history, previous incarcerations, injury to a victim, and the possession or use of a weapon. Each offense is assigned a score representing its relative severity, and each sentencing factor is assigned a score which represents the factor's relative importance in the sentencing decision.[3]

Next, to generate the sentence length range associated with each possible total score, the sentence length worksheet was used to calculate scores for individuals in the five-year cohort (1998-2003) that received a prison sentence. For each potential score within an offense category, the middle 50 percent of all sentences imposed was calculated and served as a baseline for the sentence length ranges. Then, small adjustments were made to reflect the guidelines' goals. Namely, the guidelines are designed such that drug and property offenders receive shorter sentences, while violent offenders receive slightly longer sentences (Alabama Sentencing Commission 2005).

Alabama's sentencing guidelines were initially voluntary because of concerns over constitutionality. Prior to 2004, in states with presumptive sentencing guidelines, judges could impose a sentence length above the maximum if they identified any aggravating factors, such as whether the defendant was paid to commit the offense. However, the U.S. Supreme Court, in *Blakely v. Washington*[4] and *United States v. Booker*[5], denied judges the authority to find aggravating factors. Rather, juries must establish the relevant facts required to impose a sentence above the statutory

[2]Worksheets are to be completed for the "most serious offense" at a sentencing event, which is defined as an event including "all convictions sentenced at the same time." The "most serious offense" is determined by the offense that yields the highest number of points shown on the respective Prison Sentence Length worksheet, unless a non-worksheet offense has a higher statutory maximum penalty (Alabama Sentencing Commission 2013b).

[3]Scores for each of the sentencing factors can differ across offense categories, and across disposition and sentence length worksheets within an offense category.

[4]542 U.S. 296 (2004)

[5]543 U.S. 220 (2005)

maximum. Because the ASC did not want Alabama's sentencing guidelines to be subject to constitutional scrutiny, they adopted voluntary guidelines.

The ASC expected the voluntary sentencing guidelines to be adhered to in 75 percent of the affected cases; judges would have flexibility in the remaining 25 percent to sentence outside of the suggested ranges. However, usage varied across counties; some counties did not submit any worksheets, others submitted worksheets for over 90 percent of the worksheet-applicable sentencing events. During Fiscal Year 2012 (October 2011-September 2012), counties submitted worksheets for 61 percent of worksheet-applicable sentencing events, 62.8 percent of which were considered valid (Alabama Sentencing Commission 2014).[6]

In addition to worksheet submission rates, guideline adherence is also measured by compliance with the worksheets' disposition and sentence length recommendations. For example, 36.3 percent of valid worksheets for drug offenses recommended a prison sentence, and judges adhered to the recommended "In" disposition 81.5 percent of the time. They followed the recommended "Out" disposition 73.6 percent of the time. In terms of sentence length compliance for property worksheet offenses, 42 percent of valid worksheets received a sentence within the recommended range and 35 percent received a sentence above the maximum recommended sentence length. For drug worksheet offenses, 52 percent of valid worksheets received a sentence within the recommended range and 26 percent received a sentence above the maximum recommended sentence length (Alabama Sentencing Commission 2014).[7,8]

Because worksheet adherence was so varied, the Alabama Legislature was concerned about the legal repercussions of persistent overcrowding. Namely, in response to California's

[6] Worksheets are considered valid if the conviction offense listed on the worksheet is consistent with the conviction offense found in the State Judicial Information System (SJIS) or on sentencing orders.

[7] The remaining percentage of valid worksheets is due to mixed compliance, which applies to split sentences. For most worksheets classified as mixed compliance, the incarceration portion of the split sentence fell within the recommendations, but the total sentence length exceeded the maximum recommended sentence length.

[8] Note, for all worksheet offenses, overall compliance does not appear to differ much by race: when sentencing black offenders, judges adhered to the disposition and sentence length recommendations 57.3 percent of the time, and when sentencing white offenders, judges adhered to the guidelines 61 percent of the time.

overcrowded prisons, the U.S. Supreme Court's 2011 decision in *Brown v. Plata*[9] required California to reduce its prison population within two years. Since Alabama had a history of litigation concerning prison conditions, and Alabama's prisons were more overcrowded than California's prisons, Alabama wanted to avoid a federal takeover of state institutions or a mandate requiring a reduction in the prison population. Therefore, during the 2012 Regular Session, the Alabama Legislature passed Act 2012-473, which mandated that the Initial Voluntary Sentencing Standards become presumptive for drug and property worksheet offenses, effective October 1, 2013.

As part of Act 2012-473, property offenses covered by the voluntary sentencing guidelines were divided into two subsets. The first property subset (referred to as "Property?) contains burglary offenses, for which sentencing recommendations remained voluntary. The second property subset, referred to as "Property A,? contains all other covered property offenses; worksheets for these offenses became presumptive. Sentencing recommendations for personal worksheet offenses remained voluntary.

Act 2012-473 also included additional drug offenses to be covered by the presumptive sentencing guidelines. The ASC developed scores for the new offenses by accounting for the new offenses' severity relative to the severity (and scores) of the existing offenses; the scores previously assigned to covered offenses and sentencing factors were not affected. Because a relatively more severe drug offense was added (the Unlawful Manufacture of a Controlled Substance in the first degree), the drug sentence length table had to be amended to include longer sentence length options. After these additions were made, the ASC analyzed conviction and sentencing data from a seven year time period and found that 95 percent of the prison sentences for the Unlawful Manufacture of a Controlled Substance in the first degree fell within the added sentence length ranges (Alabama Sentencing Commission 2013a).

Although judges are required to follow the disposition and sentence length recommen-

[9] 563 U.S. 493 (2011)

dations for presumptive worksheet offenses, Act 2012-473 defined aggravating and mitigating circumstances that can result in departures from the sentencing recommendations.[10] Furthermore, individuals convicted of a presumptive worksheet offense are entitled to a jury trial to establish the existence of an aggravating factor, but a jury is not required to identify the existence of a mitigating circumstance. For presumptive worksheet offenses, departures from the sentencing recommendations are subject to limited appellate review, while a judge's failure to consider the voluntary sentencing standards is not appealable (Alabama Sentencing Commission 2013b).

On December 11, 2015, the ASC adopted Act 2015-185. This "prison reform" legislation, which became effective on October 1, 2016, created a new felony offense category (Class D) and re-classified non-violent worksheet offenses as Class D offenses. All Class D felonies are covered by the presumptive sentencing guidelines, but the new category lowered the maximum sentence an individual can receive. Additional non-violent property offenses were also added to the sentencing standards.

2.4 Data

I use the National Corrections Reporting Program (NCRP), 2000 to 2016, from the National Archives of Criminal Justice Data (NACJD), to study Alabama's transition from voluntary to presumptive sentencing guidelines.[11] The NCRP provides monthly, offender-level admission and release data for individuals incarcerated in state prisons. It includes offense and offender characteristics, such as race, sex, the county where the sentence was imposed, current offense(s), total sentence length assigned by the court, date of birth, date of admission and, if applicable,

[10]The ASC defines aggravating factors as "substantial and compelling reasons justifying an exceptional sentence whereby the sentencing court may impose a departure sentence above the presumptive sentence recommendation for an offense." Mitigating factors are defined as "substantial and compelling reasons justifying an exceptional sentence whereby the sentencing court may impose a departure sentence below the presumptive sentence recommendation for an offense" (Alabama Sentencing Commission 2013b).

[11]United States. Bureau of Justice Statistics. National Corrections Reporting Program, [United States], 2000-2016. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-03-21. https://doi.org/10.3886/ICPSR37007.v1

date of release.

My sample consists of black and white males in Alabama state prisons, at least 17 years old at admission.[12] Using admissions that occurred between January 2007 and December 2015, I restrict my sample to individuals admitted for the commission of a new offense (this excludes inmates admitted for probation and parole violations).[13,14]

For my analyses, I study total sentence length, conditional on admittance into state prison. Total sentence length is the sentence length, in months, assigned by the judge for all offenses in a given sentencing event. For presumptive worksheet offenses, judges have discretion in allocating the consecutive sentencing of multiple current offenses, but the total sentence for all counts cannot exceed the maximum sentence length specified in the Prison Sentence Length Range table (Alabama Sentencing Commission 2013b). Note that in Alabama, the amount of time offenders serve in prison is likely to be shorter than the sentence length imposed by the court, so total sentence length is not equivalent to time served.[15]

In Figure 2.1, average total sentence length among both property and drug worksheet offenders appears to fall several months before the presumptive guidelines were implemented in October 2013. A possible explanation is that judges may have begun to adhere to the sentencing recommendations before the worksheets became officially presumptive. The Alabama Legislature passed the presumptive guidelines in May 2013. Through the summer and fall of 2013, members of the ASC led 30 training sessions around Alabama, particularly in counties where judges, prosecutors, and defense lawyers were less familiar with the worksheets. Because judges may have started using the presumptive guidelines prior to October 2013, I drop prison admissions that occurred between May and September 2013, which corresponds to the months in which the

[12] Alabama does not provide information on whether an inmate is Hispanic.

[13] I exclude admissions data between 2000 and 2006 because Alabama only reports a small subsample of new admissions for that period.

[14] I exclude data from October to December 2016 because, as discussed in Section 2.3, the set of offenses following the "prison reform" legislation is not necessarily comparable to the set of offenses prior to October 2016. I also exclude January to September 2016 because there appears to be some anticipatory effects, since throughout 2016, the ASC provided trainings on the guideline modifications (Alabama Sentencing Commission 2017).

[15] For example, for every 30 days served, some inmates can reduce their sentence length by 75 days.

trainings occurred.

Additionally, because of prison overcrowding, convicted offenders may wait in county jails for up to 30 days after sentencing before being transferred to state prison (Alabama Sentencing Commission 2006). Because the NCRP provides the date of admission into state prison, rather than the date of the sentencing event, I exclude admissions that occurred during October 2013 - some of the offenders recorded as entering prison in October may have been sentenced before the drug and property worksheets became presumptive. Overall, I exclude admissions occurring between May and October 2013 from all analyses, including the descriptive statistics.[16]

Table 2.1 presents descriptive statistics for all offenders admitted into Alabama state prisons for the commission of a new offense. Offenses are categorized as property, drug, burglary, personal, and other. The first four categories include both worksheet and non-worksheet offenses. The ”other” category consists of non-worksheet offenses, such as bribery or obstruction of justice, that cannot be categorized under property, drug, or personal. The descriptive statistics are divided into two periods: (1) the pre-presumptive period, January 2007 to April 2013, and (2) the post-presumptive period, November 2013 to December 2015.

Table 2.1 shows that (worksheet and non-worksheet) drug offenders make up the largest share of new admissions - over 30 percent - and property offenders make up about 20 percent of new admissions. Moreover, approximately 80 percent of all offenses are covered by the presumptive and voluntary guidelines. Property worksheet offenders make up over 87 percent of all property offense admissions, and drug worksheet offenders make up around 95 percent of all drug offense admissions.

Next, Table 2.2 presents descriptive statistics for property, drug, and personal worksheet offenders. As in Table 2.1, the descriptive statistics are divided into the pre- and post-presumptive time periods. Table 2.2 shows that black offenders make up over 40 percent of new admissions among property worksheet offenders, about 52 percent of new admissions among drug worksheet

[16] All figures include admissions occurring between May and October 2013.

offenders, and over 68 percent of new admissions among personal worksheet offenders. Offenders aged 17 to 34 make up over half of new admissions among property and drug worksheet offenders; 70 percent of personal worksheet offenders are aged 17 to 34 at the time of admission.[17]

To preface how the shift to presumptive sentencing guidelines may have affected sentence length among property and drug worksheet offenders admitted into state prison, Table 2.3 breaks down average total sentence length by race and worksheet offense category.[18] On average, black offenders receive a shorter sentence length relative to white offenders for property, drug, and personal worksheet offenses. However, among property and personal worksheet offenders, the difference in average sentence length between white and black offenders in the pre-presumptive period is not statistically significant. Furthermore, controlling for county and offense fixed effects, black offenders, on average, receive a larger sentence relative to white offenders.[19]

Table 2.3 also shows that among property and drug worksheet offenders, the difference in average sentence length between the pre- and post-presumptive time periods is statistically significant. However, among property worksheet offenders, in the post-presumptive period, the difference in average sentence length across race is not statistically significant. Among drug worksheet offenders, the difference in average sentence length across race and time periods is statistically significant. Among personal worksheet offenders, while average sentence length falls by about 19 months for black offenders between the pre- and post-presumptive periods, the differences across race and time periods are not statistically significant. The statistically significant differences across time periods serve as motivation for the following sections, which investigate the extent to which differences in average sentence length between the pre- and post-

[17]Note that the differences across offense and offender characteristics between the pre- and post-presumptive periods may be more indicative of a trend over time rather than a discrete change due to the shift to presumptive sentencing guidelines.

[18]For individuals recorded as receiving a life sentence for a drug or property offense, I replace their total sentence length with 360 months (the 99^{th} percentile of total sentence length for property and drug offenses is 240 months).

[19]Define a county as relatively lenient if the average total sentence length imposed in the county is relatively short. Within a worksheet category, there is little difference in average offense severity (or worksheet score) across relatively lenient and harsh counties. Furthermore, counties that sentence less harshly on average, admit a larger share of black offenders.

presumptive time periods can be attributed to Alabama's transition to presumptive sentencing guidelines.

2.5 Composition of Alabama's Prisons

This section describes two mechanisms through which the shift from voluntary to presumptive sentencing guidelines could affect the composition of Alabama's state prisons, namely through changes in the probability of incarceration and changes in prosecutorial discretion. While I cannot observe prosecutors' charging behavior or changes in the use of prison, this section provides suggestive evidence that the transition to presumptive guidelines affected offense and inmate composition.

Alabama's sentencing guidelines were designed to reduce the use of prison, particularly for first-time offenders. Comparing Fiscal Years 2004-2006 and Fiscal Years 2007-2008 (the shift from no sentencing guidelines to voluntary guidelines), the use of prison for all drug and property worksheet offenders declined by 5 to 11 percentage points (Alabama Sentencing Commission 2013a). Yet, there is no consensus in the literature on the effects of sentencing guidelines on court commitment rates. Some researchers find that the adoption of presumptive sentencing guidelines decreases admission rates (Nicholson-Crotty 2004; Sorensen and Stemen 2002). On the other hand, Marvell (1995) finds that sentencing guidelines did not affect court commitments; any changes to the prison population occurred through reductions in imposed sentence length.

Because Edwards, Rushin, and Colquitt (2019) finds that, compared to Alabama?s voluntary sentencing guidelines, the presumptive guidelines are more effective in altering judicial behavior, it is plausible that the probability of incarceration further declined after the sentencing guidelines became presumptive, particularly in counties where voluntary guideline usage was low.[20] A change in the use of incarceration would likely affect the type of offenders that receive a

[20]The ASC does not provide data on how the use of prison changed when the drug and property worksheets became presumptive.

prison sentence, leading to changes in inmate or offense composition and confounding the effect on average sentence length. Namely, if less severe offenders are less likely to receive prison time, then absent any changes to sentence length decisions, average total sentence length would increase. Consequently, if judges imposed shorter sentences for non-violent offenders following the shift to presumptive guidelines, the effect on average sentence length would be underestimated. Furthermore, if the changing probability of incarceration is correlated with extralegal factors, the effects of those factors on sentencing decisions in the post-presumptive period would be mis-estimated.

In terms of changes to prosecutorial discretion, according to the hydraulic displacement of discretion theory, when sentencing guidelines limit judicial discretion but place no restrictions on prosecutorial behavior, discretion is shifted to prosecutors (McCoy 1984). Typically, researchers find that following the adoption of sentencing guidelines there are little to no changes in prosecutors' charging and plea bargaining behavior, and small changes to prosecutorial discretion do not result in significant disparities (Miethe 1987; Wooldredge and Griffin 2005). On the other hand, Piehl and Bushway (2007) find evidence suggesting that sentencing guidelines give more power to prosecutors relative to judges, resulting in excessive punishment.

In Alabama, Edwards, Rushin, and Colquitt (2019) finds no evidence that prosecutors manipulated charges immediately following the shift to presumptive guidelines; there was no change in the average severity of indictment offenses. Yet, anecdotal evidence suggests that Alabama prosecutors may search for aggravating factors early on in cases so that defendants will admit to them when they plead guilty (because then a jury trial is not required to establish the existence of an aggravating factor). If this is the case, prosecutorial behavior may limit the extent to which offenders' sentence length decreases, since aggravating factors allow judges to sentence outside of the range specified in the presumptive worksheets.

Therefore, I evaluate whether offense and inmate composition changed after the presumptive sentencing guidelines were implemented. I use OLS regressions on levels of compositional outcomes and linear probability models to evaluate the following: (1) admission type (new court commitment or parole/probation violation), (2) offense composition, and (3) inmate composition, namely age at admission and race.

As described below, the shift to presumptive sentencing guidelines is associated with changes in the composition of Alabama's state prisons, and effects appear to vary across race and offense type. For example, only among white drug worksheet offenders does the probability of being admitted through a new court commitment decline in the post-presumptive period. There are also statistically significant increases in the average and median age at admission. Yet, racial composition does not appear to be affected by the shift to presumptive guidelines.

2.5.1 Admission Type

I first examine whether the composition of admission types is affected by the transition to presumptive sentencing guidelines. In particular, I focus on admissions through new court commitments and admissions through parole or probation violations. Among property and drug worksheet offenders, around 80 percent of admissions into prison are through new court commitments (see Table 2.2). Parole and probation revocations make up between 17 and 19 percent of admissions.

If the probability of incarceration declined after the guidelines became presumptive, the number of new court commitments for worksheet offenses should fall. Moreover, among property and drug worksheet offenders, the average total sentence length for offenders admitted for a parole violation is about 16 years; the average total sentence length among individuals admitted through a new court commitment is less than 5.5 years. As a result, since prosecutorial discretion is not limited by the worksheets, if prosecutors believe the guidelines to be too lenient[21], especially for individuals with a substantial criminal history, they may have an incentive to charge an individual

[21] Anecdotal evidence suggests that Alabama prosecutors believe the presumptive sentencing guidelines to be too lenient and that requiring a jury trial to prove an aggravating factor increases prosecutors' workloads (Lawson 2013; Hare 2013).

with a parole or probation violation instead of the commission of a new offense (assuming violations carry a longer sentence length).[22]

To assess the effects of the presumptive sentencing guidelines on admission type, I first estimate an OLS regression on the number of court commitment admissions and the number of parole or probation violation admissions.

$$Y_{ctr} = \beta_1^{(1)} Post_t + \beta_2^{(1)} Black_{ctr} + \beta_3^{(1)} Post_t * Black_{ctr} + \beta_4^{(1)} * t + \beta_5^{(1)} Black_{it} * t + \gamma_c^{(1)} + \gamma_q^{(1)} + \varepsilon_{ctr}^{(1)} \tag{2.1}$$

Let Y_{ctr} represent the number of court commitment admissions or parole/probation violation admissions in county c, at quarter t, and for race r. $Post_t$ is an indicator variable denoting the post-presumptive period (November 2013 to December 2015). At quarter t, for a given admit type, each county has two observations: the number of black offenders admitted to prison and the number of white offenders admitted. $Black_{ctr}$ is an indicator for the county-level observation that corresponds to black offenders. County fixed effects, $\gamma_c^{(1)}$, quarter fixed effects, $\gamma_q^{(1)}$, and separate linear time trends (at the quarterly level) for white and black offenders are included. The linear time trend is centered at 0 - the quarters in the pre-presumptive period are negative (or equal to 0 in the case of the quarter right before the transition). Quarters in the post-presumptive period are positive. Regressions are run separately by worksheet offense type, and standard errors are clustered at the county level.

The coefficients of interest in Equation 2.1 are $\beta_1^{(1)}$ and $\beta_3^{(1)}$. In regressions on the number of court commitments, if the probability of incarceration fell after the transition to presumptive sentencing guidelines, β_1 should be statistically significant, such that the number of court commitments for property and drug worksheets fell following the policy change. If the shift to presumptive guidelines did not differentially affect the number black and white offenders

[22] Prosecutors may file a parole or probation violation instead of a new case since the burden of proof is much lower in the former. The prosecutor only needs to prove that it is more likely than not that the terms of an individual's parole or probation were violated. Austin and Lawson (1998) find that in California, most technical parole violations were attributed to a new offense that was not prosecuted.

admitted, β_3 would not be statistically significant. In regressions on the number of parole or probation violations, if prosecutors charged relatively more individuals with a violation, instead of a new offense, in the post-presumptive period, β_1 should be positive. Yet, without data on prosecutors' charging behaviors, a statistically significant coefficient does not prove there were changes in how prosecutors exercised discretion.

Table 2.4 presents the coefficients from the regressions on the number of court commitments (columns labeled "New Commit") and the number of admissions for parole or probation violations (columns labeled "Violation").[23] Even-numbered columns include a coefficient on the interaction $Post_t * Black_{ctr}$; odd-numbered columns omit the interaction term.

Columns 1 to 4 list the coefficients from regressions using the sample of property worksheet offenders. For both new court commitments and parole or probation violation admissions, the coefficients on $Post_t$ and $Post_t * Black_{ctr}$ are statistically insignificant. All else equal, the shift to presumptive sentencing guidelines does not appear to have a statistically significant impact on the number of property worksheet offenders admitted for the commission of a new offense or for a parole or probation violation.

Columns 5 to 8 list the coefficients from regressions using the sample of drug worksheet offenders. In Columns 5 and 6, the coefficients on $Post_t$ are negative and statistically significant. Therefore, in Column 5, all else equal and net of time trends, the average number of court commitments for drug worksheet offenses falls by 0.614 in the post-presumptive period. In Column 6, among white offenders, all else equal and net of time trends, the average number of court commitments for drug worksheet offenses falls by 0.831 in the post-presumptive period. And while the coefficient on $Post_t * Black_{ctr}$ is not statistically significant, it is positive. A test of $\beta_1^{(1)} + \beta_3^{(1)} = 0$ suggests that for black offenders, the average number of court commitments in the post-presumptive period is not statistically different from the average number of court

[23]On average, about 21 individuals per county, each quarter, are admitted for the commission of a new property worksheet offense and 7 individuals are admitted for a parole or probation violation. For drug worksheet offenses, approximately 28 individuals per county, on average, are admitted for the commission of a new offense and 8 individuals are admitted for a parole or probation violation.

commitments in the pre-presumptive period.

For regressions on the number of admissions through parole or probation violations, in Column 7, the coefficient on $Post_t$ is positive and statistically significant. All else equal and net of time trends, the average number of admissions for a parole or probation violation increases by 0.487 after the guidelines become presumptive. In Column 8, when $Post_t * Black_{ctr}$ is included, the coefficient on $Post_t$ remains positive but becomes statistically insignificant. The coefficient on $Post_t * Black_{ctr}$ is also statistically insignificant, but a test of $\beta_1^{(1)} + \beta_3^{(1)} = 0$ suggests that among black offenders, all else equal, the average number of admissions for a parole or probation violation increases in the post-presumptive period by 0.601.[24]

Next, I estimate a linear probability model to assess whether the probability of being admitted for a new court commitment, rather than for a probation or parole violation, was affected by the transition from voluntary to presumptive sentencing guidelines.

$$\begin{aligned} Y_{ict} = {} & \beta_1^{(2)} Post_t + \beta_2^{(2)} Black_{it} + \beta_3^{(2)} Post_t * Black_{it} + \beta_4^{(2)} * t + \beta_5^{(2)} Black_{it} * t \\ & + X_{it}' \beta_6^{(2)} + \gamma_c^{(2)} + \gamma_m^{(2)} + \gamma_o^{(2)} + \varepsilon_{ict}^{(2)} \end{aligned} \tag{2.2}$$

For individual i, in county c at quarter t, Y_{ict} takes on a value of 1 if an individual is admitted for the commission of a new offense, and it takes on a value of 0 if an individual is admitted for a parole or probation violation. $Post_t$ is an indicator variable denoting the post-presumptive period. $Black_{it}$ is an indicator for whether an offender is black. Offender- and offense-level covariates, X_{it}, include the number of counts for the first offense and indicators for whether the offender was convicted of a second or third offense. County fixed effects, γ_c, month fixed effects, γ_m, offense fixed effects, γ_o, and separate linear time trends (at the quarterly level) for white and black offenders are included. Standard errors are clustered at the county level.

The coefficients of interest in Equation 2.2 are $\beta_1^{(2)}$ and $\beta_3^{(2)}$. If the probability of incar-

[24]If Equation 2.1, with the number of parole or probation violations as the dependent variable, is run separately by race, the coefficient on $Post_t$ is positive and statistically significant for both black and white drug worksheet offenders.

ceration fell after the transition to presumptive sentencing guidelines or if prosecutors began charging individuals with relatively more parole or probation violations, $\beta_1^{(2)}$ and/or $\beta_3^{(2)}$ should be statistically significant, such that the probability of being admitted into prison for the commission of a new offense, rather than a probation or parole violation, declines after the guidelines became presumptive.

Table 2.5 lists the coefficients from the regression on the probability of an individual being admitted for the commission of a new offense. In Column 1, among property worksheet offenders, the coefficient on $Post_t$ is insignificant, but the coefficient on $Post_t * Black_{it}$ is negative and statistically significant. For white offenders, all else equal, the probability of being admitted through a new court commitment does not appear to be affected by the shift to presumptive sentencing guidelines. Among black offenders, all else equal and net of time trends, the probability of being admitted for the commission of a new property worksheet offense falls by an average of 4.5 percentage points in the post-presumptive period.[25] Column 2 illustrates that among white drug worksheet offenders, all else equal, the probability of being admitted through a new court commitment fell by an average of 4.4 percentage points after the sentencing guidelines became presumptive. Yet, among black offenders, there is no statistically significant difference in the probability of a court commitment admission in the pre- and post-presumptive periods.[26,27]

Overall, while there is no statistically significant difference in the average number of property worksheet offenders admitted into prison following the shift to presumptive sentencing guidelines, the probability of a black property worksheet offender being admitted through the commission of a new offense (instead of through a parole or probation violation) decreases. Among white drug worksheet offenders, all else equal, the average number of new court commitments falls in the post-presumptive period, as does the probability of being admitted through

[25] A test of $\beta_1^{(2)} + \beta_3^{(2)} = 0$ yields a p-value of 0.0045.

[26] The coefficient on $Post_t * Black_{it}$ is positive but statistically insignificant, and a test of $\beta_1^{(2)} + \beta_3^{(2)} = 0$ yields a p-value of 0.3.

[27] All results are robust to the inclusion of an age at admission quadratic and to the exclusion of offense fixed effects and characteristics in X_{it}.

a new court commitment. Among black drug worksheet offenders, all else equal, the shift to presumptive guidelines does not appear to have a statistically significant effect on the average number of court commitments or the probability of a court commitment admission.

2.5.2 Offense Composition

Prosecutors have discretion in how to charge a defendant ? file a new case or file a parole or probation violation. They also have the authority to decide which primary offense an individual is charged with, how many counts, and how many charges (of different offenses) to bring against an individual. Therefore, to understand how the shift to presumptive sentencing guidelines may have affected the composition of offenses for which a prison sentence is imposed (possibly through changes in prosecutorial behavior), I estimate regressions on three offense composition outcomes: (1) whether an individual is admitted for a property or drug worksheet offense relative to all other offenses, (2) the severity of admitted worksheet offenses[28], and (3) whether a presumptive worksheet offender enters prison with a second or third offense.

First, to study how the probability an individual is admitted for a property or drug worksheet offense, relative to all other offenses, changes after the guidelines became presumptive, I estimate the linear probability model in Equation 2.2 (excluding offense fixed effects).[29] I estimate Equation 2.2 using two samples. In the first sample, the outcome variable takes on a value of 1 if an offender is admitted for a property worksheet offense and 0 otherwise. In the second sample, the outcome variable takes on a value of 1 if an offender is admitted for a drug worksheet offense and 0 otherwise.

Columns 1 and 2 of Table 2.6 present the coefficients from the estimation of this model. Recall that in the prior section, there was no statistically significant difference in the average number of property worksheet offenders admitted after the guidelines became presumptive

[28]Offense severity is measured using the score assigned to an offense in the Prison Sentence Length worksheet.
[29]From this point onward, the sample is restricted to new court commitments.

(Columns 1 and 2 of Table 2.4). But, there was a statistically significant decline in the average number of white drug worksheet offenders admitted (Columns 5 and 6 of Table 2.4). In Column 1 of Table 2.6, all else equal and net of time trends, there is no statistically significant difference in the probability that an offender is admitted for a property worksheet offense in the post-presumptive period. In Column 2, all else equal and net of time trends, the probability a white offender is admitted for a drug worksheet offense, falls by 3.1 percentage points in the post-presumptive period. The coefficient on $Post_t * Black_{it}$ is positive and statistically significant, and a test of $\beta_1^{(2)} + \beta_3^{(2)} = 0$ is statistically insignificant. Therefore, among black offenders, all else equal, there is no statistically significant change in the probability that following the policy change.[30]

Next, I estimate whether average offense severity, conditional on incarceration, may have been affected by the transition to presumptive guidelines (possibly through a reduction in the use of prison). Holding constant prosecutors charging behavior, if individuals convicted of less severe worksheet offenses are less likely to be incarcerated following the policy change, the average severity of worksheet offenses among incarcerated offenders will increase. On the other hand, it might be the case that prosecutors are changing their charging behavior for only the most severe worksheet offenses, such that they start charging offenders with relatively more non-worksheet offenses in an attempt to increase sentence length if an individual is convicted.[31] Then, the average severity among incarcerated worksheet offenders will decline. The descriptive statistics in Columns 1 and 2 of Table 2.2 show that among property worksheet offenders, there is little change in average offense severity across the pre- and post-presumptive periods. Among

[30] All results are robust to the inclusion of an age at admission quadratic and to the exclusion of the characteristics in X_{it}.

[31] The probability an individual is admitted for a presumptive worksheet offense may also be affected by the transition to presumptive guidelines if prosecutors substitute worksheet offense charges for non-worksheet charges. For example, an individual may be charged with possession of chemicals used in the manufacturing of drugs instead of the applicable drug worksheet offense (Kirby 2013). There does not appear to be a statistically significant shift from worksheet to non-worksheet offenses within a given offense. Yet, since property and drug worksheet offenses make up at least 87 percent of all property and drug offenses, the sample size of property and drug non-worksheet offenses is too small to detect any meaningful results.

drug worksheet offenders, Columns 3 and 4 display a statistically significant increase in average offense severity.

Again, I estimate Equation 2.2, this time using a continuous measure of offense severity as the dependent variable. Offense severity is measured by the score each offense receives on the Prison Sentence Length worksheets. I use the same covariates in X_{it} but omit offense fixed effects. Columns 3 and 4 of Table 2.6 list the coefficients from the regression on presumptive worksheet offense severity. In Column 3, among property worksheet offenders, the coefficient on $Post_t$ is negative but statistically insignificant. While the coefficient on $Post_t * Black_{it}$ is positive and statistically significant, a test of $\beta_1^{(2)} + \beta_3^{(2)} = 0$ is statistically insignificant (the p-value is 0.12). Therefore, there is no statistically significant difference in average offense severity following the shift to presumptive sentencing guidelines. In Column 4, among white drug worksheet offenders, all else equal and net of time trends, average severity falls by 7.15 points in post-presumptive period. The coefficient on $Post_t * Black_{it}$ is statistically significant and positive. As a result, among black offenders, there is no statistically significant effect on average offense severity after the guidelines became presumptive.[32]

Lastly, because prosecutors also have the discretion to bring multiple charges, I estimate how the adoption of the presumptive sentencing guidelines may affect whether an individual is more likely to enter prison with a second or third offense.[33] The more offenses an individual is convicted of in the same sentencing event, the higher the worksheet score, possibly increasing the minimum and maximum sentence mandated by the guidelines. The descriptive statistics in Table 2.2 show that about 30 percent of property worksheet offenders are incarcerated with a second offense and about 20 percent are incarcerated with a second and third offense. Among drug worksheet offenders, about 24 percent are incarcerated with a second offense and about 8

[32] Edwards, Rushin, and Colquitt (2019) finds that the average severity of the indictment offense did not change in the weeks following the implementation of the presumptive guidelines.

[33] Alschuler (1968) labels this behavior horizontal overcharging. See Miethe (1987) for a discussion of prosecutors bringing multiple charges against a defendant in the context of Minnesota's presumptive sentencing guidelines in footnote 26.

percent are incarcerated with a second and third offense.

I estimate the probability of a presumptive worksheet offender entering prison with a second or third offense using Equation 2.2 (excluding the indicators, in X_{it}, for whether an offender was convicted of a second or third offense). The dependent variable takes on a value of 1 if an offender was convicted for a second or third offense. Columns 5 and 6 of Table 2.6 present the coefficients from this regression. In Column 5, among white property worksheet offenders, the probability of being admitted into prison with a second or third offense falls by 4.5 percentage points in the post-presumptive period, all else equal. The coefficient on $Post_t * Black_{it}$ is not statistically significant but is positive. A test of $\beta_1^{(2)} + \beta_3^{(2)} = 0$ is statistically insignificant - all else equal, black property worksheet offenders' probability of entering prison with a second or third offense is not statistically different across the pre- and post-presumptive periods. In Column 8, among drug worksheet offenders, the shift to presumptive guidelines is not associated with a statistically significant change in the probability of being admitted with a second or third offense.[34]

I also run regressions on the number of counts for the first offense (results not provided), since the worksheet score is increasing in the number of counts. Only for black drug worksheet offenders does the average number of counts change (decrease) in the post-presumptive period.

Overall, the shift from voluntary to presumptive sentencing guidelines appears to be associated with some changes in offense composition as measured by the probability an individual is admitted for a presumptive worksheet offense, the average offense severity, and the probability an individual enters prison with a second or third offense. Among white drug offenders, following the policy change, there is a decline in the probability an individual is admitted for a drug worksheet offense and a decline in the average severity of admitted offenses. And among white property worksheet offenders, the probability of entering prison with a second or third offense

[34] All results are robust to the inclusion of an age at admission quadratic. When I exclude offense fixed effects and the number of counts for the first offense, the coefficients on $Post_t$ and $Post_t * Black_{it}$, in Column 8, remain statistically insignificant but reverse sign.

falls. Yet, without more information on changes in the use of prison or in prosecutors' behavior, I cannot determine the mechanism through which offense composition is affected.

2.5.3 Inmate Demographics

In the NCRP, Alabama only reports two demographic variables: age at admission and race. Because the transition to presumptive sentencing guidelines is likely to have reduced the probability of incarceration, particularly for first-time offenders, average age at admission may increase (assuming age is positively associated with prior offending). Furthermore, if, for example, first-time offenders admitted into prison in the pre-presumptive period are disproportionately black, then the share of black offenders may fall as the probability of incarceration declines. Similarly, the share of black offenders might decline if black offenders are disproportionately incarcerated for less severe offenses in the pre-presumptive period (assuming these less severe offenses are less likely to receive a prison sentence in the post-presumptive period).

2.5.3.1 Age at Admission

Figure 2.3 shows how the average and median age (at the time of admission) have evolved over time. Quarter 0 denotes the quarter before the presumptive sentencing guidelines were implemented. The solid black line corresponds to white offenders and the gray solid line corresponds to black offenders. Figures 2.3a and 2.3b illustrate the evolution of the average and median ages at admission among property worksheet offenders, respectively. They show that the average white offender is usually younger than the average black offender. Additionally, they illustrate that following the shift to presumptive guidelines, there appears to be an increase in the average and median ages for both black and white offenders.

Figures 2.3c and 2.3d present a time-series of the average and median ages at admission among drug worksheet offenders, respectively. They show that among drug worksheet offenders, black offenders are younger on average and at the median. Furthermore, there appears to be an increase in the average age of black offenders. However, this increase seems to occur in the quarter before the guidelines became presumptive. In Figure 2.3d, the changes to median age at admission are more evident, but as with average age, the median age for black offenders increases the quarter before the transition.

To estimate how the shift to presumptive guidelines may have affected the average age at admission, I estimate Equation 2.2, where Y_{ict} is individual $i's$ age at admission (in county c, at quarter t). I also run a median regression on age at admission, that is otherwise identical. The coefficients of interest are $\beta_1^{(2)}$ and $\beta_3^{(2)}$.

Table 2.7 presents the coefficients from regressions on age at admission using the sample of property worksheet offenders. The first three columns list coefficients from the OLS regressions, and the last three columns list coefficients from the median regressions. Columns 1 and 4 estimate regressions on age at admission using the sample of all property worksheet offenders. In Column 1, the coefficient on $Post_t$ is statistically significant; on average, white property worksheet offenders are, all else equal, about 0.887 years older in the post-presumptive period. While the coefficient on $Post_t * Black_{it}$ is not statistically significant, a test of $\beta_1^{(2)} + \beta_3^{(2)} = 0$ suggests that, on average, black property worksheet offenders are, all else equal, about 1.21 years older in the post-presumptive period. Columns 2 and 3, which list coefficients estimated separately by race, also support these conclusions. As with average age, Column 4 indicates that, following the shift to presumptive guidelines, the median age at admission among white offenders increased by about 1.378 years. Although the coefficient on $Post_t * Black_{it}$ is not statistically significant, a test of $\beta_1^{(2)} + \beta_3^{(2)} = 0$ suggests that, at the median, black property worksheet offenders are, all else equal, about 1.769 years older in the post-presumptive period. The coefficients on $Post_t$ in Columns 5 and 6 are positive and statistically significant as well.

Table 2.8 presents the coefficients from regressions on age at admission using the sample of drug worksheet offenders. The first three columns list coefficients from the OLS regressions, and the last three columns list coefficients from the median regressions. Columns 1 and 4 estimate regressions on age at admission using the sample of all drug worksheet offenders. Column 1 shows that among drug worksheet offenders, the coefficients on $Post_t$ and $Post_t * Black_{it}$ are statistically insignificant - all else equal, the average age at admission for drug offenders is not statistically different across the pre- and post-presumptive periods.[35] In Columns 2 and 3, the coefficients on $Post_t$, in regressions estimated separately by race, are also insignificant. In Column 4, the coefficients on $Post_t$ and $Post_t * Black_{it}$ are not statistically significant. However, a test of $\beta_1^{(2)} + \beta_3^{(2)} = 0$ suggests that among black drug worksheet offenders, the transition to presumptive guidelines is associated with a 0.63 increase in median age at admission. Similarly, the coefficient on $Post_t$ in Column 5, which uses the sample of black drug worksheet offenders, is positive and statistically significant. In Column 6, which uses the sample of white drug worksheet offenders, the coefficient on $Post_t$ is statistically insignificant.[36]

2.5.3.2 Race

Since 60 percent of drug worksheet offenders aged 17 to 34 (at the time of admission) are black, if the use of prison for young, first-time offenders fell, then the share of black drug worksheet offenders may fall after the transition to presumptive guidelines. Similarly, average severity among black drug worksheet offenders is less than that of white offenders. So if fewer, less severe worksheet offenses result in a prison sentence, the share of drug worksheet offenders that are black may decline in the post-presumptive period.

Figure 2.4 illustrates how the share of black offenders has changed over time. The black solid line represents the share of black offenders among property worksheet offenders. The gray solid line represents the share of black offenders among drug worksheet offenders. Quarter 0 represents the quarter before the sentencing guidelines became presumptive. In Figure 2.4, there is no evident discrete change in the share of black offenders following the shift to presumptive

[35] A test of $\beta_1^{(2)} + \beta_3^{(2)} = 0$ is also statistically insignificant.

[36] The results among property worksheet offenders are robust to the exclusion of characteristics in X_{it} and offense fixed affects. Among drug worksheet offenders, excluding these variables from the OLS regressions results in statistically significant (and positive) coefficients on $Post_t$.

guidelines.

More formally, I estimate a linear probability model, estimating the probability that an incarcerated offender is black:

$$Y_{ict} = \beta_1^{(3)} Post_t + X_{it}' \beta_2^{(3)} + \beta_3^{(3)} * t + \gamma_c^{(3)} + \gamma_m^{(3)} + \gamma_o^{(3)} + \varepsilon_{ict}^{(3)} \tag{2.3}$$

Let Y_{ict} be an indicator for whether individual i, in county c at quarter t, is black. Regressions are run separately among property and drug worksheet offenders. $Post_t$ is an indicator variable denoting the post-presumptive period. Offender- and offense-level covariates, X_{it}, include the number of counts for the first offense and indicators for whether the offender was convicted of a second or third offense, controlling for the offense type of the additional convictions. County fixed effects, $\gamma_c^{(3)}$, month fixed effects, $\gamma_m^{(3)}$, offense fixed effects, $\gamma_o^{(3)}$, and a linear time trend (at the quarterly level) are included. Standard errors are clustered at the county level. The coefficient of interest in Equation 2.3 is $\beta_1^{(3)}$.

Table 2.9 presents the coefficients from the estimation of Equation 2.3, the probability an incarcerated property or drug worksheet offender is black. In Columns 1 and 2, the coefficients on post are statistically insignificant.[37] Therefore, while there are statistically significant increases in the average and median age at admission, particularly among black worksheet offenders, there are no statistically significant changes in racial composition associated with the transition to presumptive sentencing guidelines.

2.6 Total Sentence Length

Alabama's sentencing guidelines are intended to reduce sentence lengths for a subset of property and drug offenders. Yet, conditional on incarceration, the expected effect on total

[37] I also run Equation 2.3 separately for each age group: 17-34, 35-49, and 50+. Conditional on age group, there is no statistically significant difference in the probability of an offender being black across the pre- and post-presumptive periods.

sentence length from shifting to presumptive guidelines is ambiguous. The change in sentence length could, instead, be driven by changes in the types of offenders receiving prison sentences. For example, if the probability of incarceration changed, such that offenders that received shorter sentences in the pre-presumptive period are more likely to receive an alternative sanction in the post-presumptive period, then average sentence length would increase. Note that, at the time the guidelines became presumptive, there were no other policy changes in Alabama that would affect sentencing decisions.

As a starting point, I run a difference regression assessing how sentence length may have changed following the transition to presumptive sentencing guidelines. Then, I estimate a difference-in-differences model using personal worksheet offenses as a control group. In both models, I also study whether the shift to presumptive sentencing guidelines had different impacts on white and black offenders' sentence lengths. Because the NCRP does not provide data on an offender's criminal history, any differences across race do not imply racial disparities in sentence length. Instead, if race is correlated with legal factors, differences in sentence length outcomes may be due to to prior offending (Engen and Gainey 2000).

Overall, I find that following the adoption of the presumptive sentencing guidelines, there was a statistically significant (percent) decline in average total sentence length among property worksheet offenders and among white drug worksheet offenders. While the difference estimates also indicate a statistically significant decline among black drug worksheet offenders, the difference-in-differences estimates suggest that, relative to black personal worksheet offenders, there was no statistically significant difference in the percent change in sentence length.

2.6.1 Difference

2.6.1.1 Methodology

First, I estimate an OLS regression on log total sentence length. The estimating equation is:

$$\begin{aligned} Y_{ict} = {} & \beta_1^{(4)} Post_t + \beta_2^{(4)} Black_{it} + \beta_3^{(4)} Post_t * Black_{it} + \beta_4^{(4)} * t + \beta_5^{(4)} Black_{it} * t + \beta_6^{(4)} Post_t * t \\ & + X_{it}' \beta_7^{(4)} + \gamma_c^{(4)} + \gamma_m^{(4)} + \gamma_o^{(4)} + \varepsilon_{ict}^{(4)} \end{aligned} \tag{2.4}$$

Let Y_{ict} represent the log total sentence length assigned to individual i, in county c, in quarter t. $Post_t$ is an indicator variable denoting the post-presumptive period. $Black_{it}$ is an indicator for whether an offender is black. Offender- and offense-level covariates, X_{it}, include the number of counts for the first offense and indicators for whether the offender was convicted of a second or third offense, controlling for the offense type of the additional convictions.[38] County fixed effects, $\gamma_c^{(4)}$, month fixed effects, $\gamma_m^{(4)}$, offense fixed effects $\gamma_o^{(4)}$, and separate linear time trends (at the quarterly level) for white and black offenders are included. Standard errors are clustered at the county level. Regressions are run separately by worksheet offense type.

The coefficients of interest in Equation 2.4 are: (1) $\beta_1^{(4)}$ estimates the percent change in sentence length associated with the shift to presumptive sentencing guidelines, (2) $\beta_3^{(4)}$ estimates whether the transition to presumptive guidelines impacted the percent change in sentence length differently across race, and (3) $\beta_6^{(4)}$ estimates whether the slope of the linear time trends differed across the pre- and post-presumptive periods.

[38] Because average age at admission seems to increase in the post-presumptive period, I omit the age at admission quadratic from X_{it}.

2.6.1.2 Results

Figure 2.5 graphs the residuals, averaged within a quarter, from a regression on log total sentence length, controlling for race and county, month, and offense fixed effects.[39] Regressions are run separately for property and drug worksheet offenders. The black circle markers represent average residuals in the pre-presumptive period; the gray triangle markers represent average residuals in the post-presumptive period. Linear fit lines, estimated separately for the pre- and post-presumptive periods, are plotted. In Figure 2.5a, which graphs the averaged residuals from a regression on property worksheet offenders, there is a discrete decline around the time of the transition to the presumptive guidelines. The decline appears to start prior to October 2013 because, as discussed in Section 2.4, the ASC had led guideline trainings between May and September 2013. As a result, judges may have begun to use the presumptive guidelines before they had been officially implemented in October of 2013.

The graph, in Figure 2.5b, of the averaged residuals from a regression on drug worksheet offenders illustrate that, in addition to a discrete drop around the time of the transition, average sentence length may decline at a faster rate in the post-presumptive period. However, it is also possible that the change in slope of the linear time trend is a continuation of a trend that changed about 10 quarters prior to the implementation of the presumptive guidelines.[40]

More formally, Tables 2.10 and 2.11 present the coefficients from the OLS regressions on log total sentence length among property and drug worksheet offenders, respectively. All regressions include the covariates in X_{it} and offense fixed effects. Column 1 regresses log total sentence length without county and month fixed effects and no interaction terms. Column 2 introduces county and month fixed effects. Column 3 adds in $Post_t * Black_{it}$ and $Post_t * t$. Columns 4 and 5 include separate linear time trends for white and black offenders.

[39] Admissions occurring between May and October 2013 are included.

[40] Figure 2.1b shows that average total sentence length for black drug worksheet offenders may have started declining about ten quarters prior to the transition to presumptive guidelines - the gap between average sentence length for black and white offenders increases prior to the transition.

In Table 2.10, among property worksheet offenders, the coefficient on $Black_{it}$ is only statistically significant in Columns 2 and 3, which do not include separate time trends for white and black offenders. In Columns 4 and 5, when separate time trends by race are included, the time trend for black offenders is statistically significant. Recall that the linear time trend is centered at 0 - the quarters in the pre-presumptive period are negative (or equal to 0 in the case of the quarter right before the transition). Quarters in the post-presumptive period are positive. Therefore, all else equal, Columns 2 through 5 indicate that in the pre-presumptive period, relative to white offenders, black property worksheet offenders had longer average sentence lengths.

Furthermore, in all columns of Table 2.10, the coefficient on $Post_t$ is negative and statistically significant. In Column 3, the coefficient on $Post_t * Black_{it}$ is negative and statistically significant, suggesting that, all else equal and net of the time trend, relative to white offenders in the post-presumptive period, average sentence length for black property worksheet offenders fell an additional 9.9 percent. Yet, once a separate time trend for black and white offenders is included, the coefficient on $Post_t * Black_{it}$ becomes insignificant. Therefore, in Column 4, all else equal and net of time trends, the average sentence length for black and white offenders falls by approximately 17.2 percent (or 11.3 months[41]) in the post-presumptive period.

Table 2.11 suggests a similar story among drug worksheet offenders. The coefficient on $Black_{it}$ is statistically significant in Column 3, but becomes insignificant when separate time trends for black and white offenders are included (Columns 4 and 5). Nonetheless, both Columns 3 and 5 suggest that, all else equal, in the pre-presumptive period, black offenders had longer average sentence lengths relative to white offenders. In all but Column 3, the coefficient on $Post_t$ is negative and statistically significant. In Columns 3 and 4, the coefficient on $Post_t * Black_{it}$ is statistically insignificant, but a test of $\beta_1^{(4)} + \beta_3^{(4)} = 0$ indicates that, among black offenders, all else equal, the shift to presumptive guidelines was associated with a 12.5 to 13.3 percent

[41] I estimate the effect in months by multiplying $Post_t$ and average sentence length (among both black and white offenders) in the pre-presumptive period.

decrease in average sentence length (or over 8 months).[42] In Column 4, among white offenders, average sentence length fell by 14.3 percent (about 9.5 months). The coefficients on $Post_t * t$, in Columns 3 and 5, are also statistically insignificant. Column 5 suggests that following the shift to presumptive guidelines, all else equal and net of time trends, average sentence length for drug worksheet offenders fell by 8.7 percent (or 5.8 months).[43]

Overall, I find that following the adoption of the presumptive sentencing guidelines, average sentence length for property and drug worksheet offenders falls, as the guidelines had intended. All else equal, accounting for a difference in time trends across race, the transition does not appear to affect black and white offenders differently.

2.6.2 Difference-in-Differences

2.6.2.1 Methodology

The transition from voluntary to presumptive sentencing guidelines in Alabama provides a quasi-experimental setting where a difference-in-differences approach can be used. I use personal worksheet offenses as the control group.[44]

I estimate the following difference-in-differences model:

$$\begin{aligned} Y_{ict} =& \beta_1^{(5)} Post_t + \beta_2^{(5)} Black_{it} + \beta_3^{(5)} * t \\ & + \beta_4^{(5)} Post_t * Black_{it} + \beta_5^{(5)} Post_t * WS_{it} + \beta_6^{(5)} Black_{it} * WS_{it} + \beta_7^{(5)} Black_{it} * t \\ & + \beta_8^{(5)} Post_t * Black_{it} * WS_{it} + X_{it}' \beta_9^{(5)} + \gamma_c^{(5)} + \gamma_m^{(5)} + \gamma_o^{(5)} + \varepsilon_{ict}^{(5)} \end{aligned} \tag{2.5}$$

[42]The p-values on $\beta_1^{(4)} + \beta_3^{(4)} = 0$ (from Columns 3 and 5) are 0.003 and 0.02, respectively.

[43]When you include an age at admission quadratic in Equation 2.4, the coefficient on $Post_t * Black_{it}$, in Column 3, is negative and statistically significant. It becomes insignificant when separate time trends are included for white and black offenders. Furthermore, among drug worksheet offenders, the coefficient on $Post_t * t$ is statistically significant when age at admission is included, but the sign on the coefficient is unchanged.

[44]The shift to presumptive guidelines may have affected burglary worksheet offenses, even though those worksheets remained voluntary. Relative to the pre-presumptive period, average total sentence length among burglary worksheet offenders seems to increase in the post-presumptive period. Additionally, I do not use "other" offenses and non-worksheet personal offenses because in some regression specifications, changes in black offenders' sentence length is associated with the transition to presumptive guidelines.

Let Y_{ict} represent the log total sentence length assigned to individual i, in county c, in quarter t. $Post_t$ is an indicator variable denoting the post-presumptive period. $Black_{it}$ is an indicator for whether an offender is black. WS_{it} is an indicator for whether an individual is admitted for a presumptive worksheet offense.[45] Offender- and offense-level covariates, X_{it}, include the number of counts for the first offense and indicators for whether the offender was convicted of a second or third offense, controlling for the offense type of the additional convictions.[46] County fixed effects, $\gamma_c^{(5)}$, month fixed effects, $\gamma_m^{(5)}$, offense fixed effects $\gamma_o^{(5)}$, and separate linear time trends (at the quarterly level) for white and black offenders are included. Standard errors are clustered at the county level. Regressions are run separately by worksheet offense type.

In Equation (2.5), the coefficients of interest are: (1) $\beta_5^{(5)}$, which estimates the average change in log total sentence length between presumptive and voluntary (personal) worksheet offenses before and after the transition to presumptive sentencing guidelines and (2) $\beta_8^{(5)}$, which, controlling for prison-wide trends, estimates whether the transition affected the percent change in sentence length differently for black and white presumptive worksheet offenders.

2.6.2.2 Results

Table 2.12 displays the coefficients from the estimation of Equation 2.5. Coefficients in Columns 1 to 3 are estimated using the sample of property and personal worksheet offenders, and coefficients in Columns 4 to 6 are estimated using the sample of drug and personal worksheet offenders. Columns 1 and 4 do not include separate time trends for black and white offenders. Columns 2 and 5 exclude the coefficient on $Post_t * Black_{it} * WS_{it}$ but include separate linear time trends for white and black offenders. Columns 3 and 6 include both the coefficient on $Post_t * Black_{it} * WS_{it}$ and separate linear time trends for black and white offenders.

Among property worksheet offenders, Columns 1 through 3 of Table 2.12 show that the

[45] I do not include a term for WS_{it} in Equation 2.5 because I include offense level fixed effects.

[46] As in Section 2.6.1.1, because average age at admission seems to increase in the post-presumptive period, I omit the age at admission quadratic from X_{it}.

coefficients on $Post_t * WS_{it}$ are statistically significant. In Columns 1 and 3, the coefficients on $Post_t * Black_{it} * WS_{it}$ are not statistically significant. Therefore, estimates in Column 3 suggest that following the transition to presumptive sentencing guidelines, sentence length fell by about 18 percent (or over 12 months) for white and black property worksheet offenders.[47] Relative to the percent change in sentence length among personal worksheet offenders, sentence length among property worksheet offenders fell by 23 percent (or 15 months). Additionally, controlling for prison-wide trends in sentence length, all else equal, the difference in the percent change in total sentence length between white and black property worksheet offenders is not statistically significant. Note that the magnitude of the OLS estimate on $Post_t$ (-0.156), in Column 4 of Table 2.10, is less than the magnitude of the difference-in-differences point estimate on $Post_t * WS_{it}$ (-0.233), in Column 3 of Table 2.12.

Similarly, among drug worksheet offenders, Columns 4 through 6 of Table 2.12 show that the coefficients on $Post_t * WS_{it}$ are statistically significant. The coefficients on $Post_t * Black_{it} * WS_{it}$ are positive but not statistically significant. Therefore, estimates in Column 6 suggest that following the transition to presumptive sentencing guidelines, sentence length fell over 12.4 percent (or about 8 months) for white and black drug worksheet offenders.[48] Relative to the percent change in sentence length among white personal worksheet offenders, sentence length among white drug worksheet offenders fell 11.8 percent. Yet, there is no statistically significant difference in the percent change in sentence length between black personal and drug worksheet offenders.[49] Additionally, controlling for prison-wide trends in sentence length, all else equal, the difference in the percent change in total sentence length between white and black drug worksheet offenders is not statistically significant. Note that the magnitude of the OLS estimate on $Post_t$ (-0.143), in Column 4 of Table 2.11, is greater than the magnitude of the difference-in-differences point estimate on $Post_t * WS_{it}$ (-0.118), in Column 6 of Table 2.12.

[47] Joint tests of significance on $\beta_1^{(5)} + \beta_5^{(5)} = 0$ and $\beta_1^{(5)} + \beta_4^{(5)} + \beta_5^{(5)} + \beta_8^{(5)} = 0$ statistically significant.

[48] Joint tests of significance on $\beta_1^{(5)} + \beta_5^{(5)} = 0$ and $\beta_1^{(5)} + \beta_4^{(5)} + \beta_5^{(5)} + \beta_8^{(5)} = 0$ are statistically significant.

[49] A joint test of significance on $\beta_5^{(5)} + \beta_8^{(5)} = 0$ is insignificant.

Therefore, using personal worksheet offenders as a control group, the shift to presumptive sentencing guidelines led to a statistically significant decline in the average total sentence length among property worksheet offenders, all else equal. As in Section 2.6.1, there are no statistically significant differences in the percent change in total sentence length across race (controlling for prison-wide trends). In contrast, relative to white personal worksheet offenders, the percent decline in sentence length among white drug worksheet offenders is relatively larger. But, following the shift to presumptive sentencing guidelines, there is no statistically significant difference in the percent change in sentence length between black personal and drug worksheet offenders.

Relative to the results found in Edwards, Rushin, and Colquitt (2019), my difference-in-differences estimates are smaller in magnitude (a decline between 30 and 41 percent relative to a decline between 11 and 23 percent, respectively). Because Edwards, Rushin, and Colquitt (2019) combine the disposition and sentence length decisions into one measure, this suggests that in addition to a decline in sentence length following the shift from voluntary to presumptive sentencing guidelines, there may have been a reduction in the number of offenders receiving a prison sentence.

2.6.2.3 Robustness Checks

This section discusses the results from several robustness analyses. In general, personal worksheet offenders appear to be a suitable control group. However, there are instances, prior to 2011, where the trends in log total sentence length between presumptive and personal worksheet offenders may be systematically deviating from one another. While I would caution against interpreting the difference-in-differences estimates as causal, there is strong evidence that, conditional on incarceration, the shift from voluntary to presumptive sentencing guidelines decreased average total sentence length for property and drug worksheet offenders.

First, I investigate the parallel trends assumption. The identifying assumption for the difference-in-differences approach is that, in the absence of the shift to presumptive guidelines, the change in log total sentence length among presumptive worksheet offenders before and after the transition would have been the same as the change among personal worksheet offenders. Otherwise, the estimated effects on sentence length may reflect differential time trends between presumptive and personal worksheet offenders, and not changes resulting from the shift in guidelines.

Figure 2.6 shows the evolution of average log total sentence length among presumptive worksheet offenders and personal worksheet offenders. Figures are presented separately by race. The gray solid line represents average log sentence length among personal worksheet offenders. In Figures 2.6a and 2.6b, the black solid line represents the average log sentence length among property worksheet offenders; in Figures 2.6c and 2.6d it represents average log sentence length among drug worksheet offenders. In Figures 2.6a, 2.6b, and 2.6c, despite variation in log total sentence length from quarter to quarter, the trends appear parallel. However, in Figure 2.6d, average log sentence length among black drug worksheet offenders begins to decline around April 2012, violating the parallel trends assumption.

More formally, I test whether the time trends in log total sentence length among personal and presumptive worksheet offenders were statistically different from one another in the pre-presumptive period. Restricting the sample to pre-presumptive admissions, and estimating Equation 2.5 (excluding the $Post_t$ coefficients) separately among property and drug worksheet offenders, I do not find a statistically significant difference in the worksheet-specific linear time trends.[50] I also estimate a regression with worksheet-specific quarter dummies (leads and lags). While all coefficients on the worksheet-specific (pre-presumptive) quarter dummies are statistically insignificant, their 95 percent confidence intervals are very large. The estimates cannot rule out effects as large as 50 percent among presumptive worksheet offenders.

Next, I perform three placebo tests. In the first placebo test, I restrict my sample to

[50]This conclusion is robust to the inclusion of different sets of control variables and is not affected by whether I use separate time trends for black and white offenders.

personal worksheet offenders and randomly assign "treatment" status. Then, I simulate the p-values on the coefficients of interest from Equation 2.5.[51] Following the transition from voluntary to presumptive sentencing guidelines, I find no statistically significant differences in the change in sentence length across the "treatment" and control groups.[52]

In the second placebo test, I again estimate Equation 2.5, comparing the effects on log total sentence length across "other" offenses and personal worksheet offenses. When I include a single time trend, all coefficients of interest (equivalent to Columns 1 and 4 in Table 2.12) are statistically insignificant. But a test of $\beta_4^{(5)} + \beta_8^{(5)}$ is statistically significant - following the transition to presumptive guidelines, relative to white "other" offenders, there is a statistically significant change in log total sentence length among black "other" offenders. In the iterations that contain a separate time trend by race (equivalent to Columns 3 and 5 in Table 2.12), all else equal, there are no statistically significant differences in changes to log sentence length between the alternative control group ("other" offenses) and the original control group (personal worksheet offenses).[53]

Lastly, I estimate Equation 2.5, changing the date of the policy change. I exclude admissions that occurred after May 2013 (the post-presumptive period) and run 20 regressions using each quarter in the pre-presumptive period as the start of the placebo "post-presumptive" period. When estimating the differences-in-differences model using the sample of drug and personal worksheet offenders, no coefficients on $Post_t$ or on interactions with $Post_t$ are statistically significant. When performing the joint tests of significance, I find that following the placebo "post-presumptive" period, black drug worksheet offenders have statistically significant changes in log total sentence length in three quarters of 2010. And, all else equal, the difference in the change in log total sentence length between white and black drug worksheet offenders is statistically significant in two quarters of 2010.

[51] I ran a simulation of 1000 repetitions and bootstrapped the standard errors on the p-values. The seed is 1234.

[52] Results are robust to different sets of controls.

[53] All results are generally robust to different sets of controls. If I only control for county, month, and offense fixed effects, the coefficient on $Post_t * Black_{it}$ is statistically significant.

When estimating the differences-in-differences model using the sample property and personal worksheet offenders, the results are less encouraging. The coefficient on $Post_t * WS_{it}$ and $Post_t * Black_{it} * WS_{it}$ are statistically significant when the placebo "post-presumptive" periods are designated as either of the first two quarters of 2008. The test of joint significance on $\beta_4^{(5)} + \beta_4^{(5)} + \beta_5^{(5)} + \beta_8^{(5)} = 0$ is statistically significant in six instances between October 2007 and April 2009. Therefore, in these instances, black property worksheet offenders had a statistically significant change in log total sentence length following the placebo "post-presumptive" period. Moreover, the joint test of $\beta_4^{(5)} + \beta_4^{(5)} = 0$ is statistically significant in 10 instances (each quarter between October 2007 and April 2009 and each quarter between October 2010 and April 2011). Namely, following the placebo policy change, all else equal and controlling for prison-wide trends, the difference in the change in log total sentence length between white and black property worksheet offenders is statistically significant.

Given these robustness checks, personal worksheet offenders are, generally, a suitable control group. However, because of the possible deviations in time trends prior to 2011, the difference-in-differences estimates should be interpreted as suggestive, rather than causal.

2.7 Quantile Regression

Alabama's sentencing guidelines are based on historical sentencing practices, with small adjustments made so that non-violent offenders sentenced to prison receive slightly shorter sentences than they would have received in the past. Furthermore, Edwards, Rushin, and Colquitt (2019) shows that Alabama's shift to presumptive sentencing guidelines forced the judges that sentenced most harshly (pre-guidelines) to reduce their imposed sentence lengths the most. Therefore, conditional on incarceration, rather than the shift to presumptive guidelines having a constant effect across the distribution of sentence length decisions, the largest declines in sentence length may occur in the right tail of the sentence length distribution.

To illustrate how the distribution of sentence length has evolved over time, for each fiscal year between 2007 and 2015, Figures 2.7 and 2.8 graph box plots of the total sentence length imposed on offenders incarcerated for property and drug worksheet offenses, respectively.[54,55] Within each worksheet offense category, I provide separate box plots for white and black offenders. However, the box plots do not account for other offender and offense characteristics, so it is possible that any differences in distributional patterns across race disappear once these factors are controlled for. The length of each gray box represents the inter-quartile range, and the vertical line in the gray box denotes median sentence length. The whiskers represent 1.5 times the inter-quartile range, and the circle markers represent potential outliers.[56]

Figure 2.7 shows how the distribution of sentence length among individuals admitted for property worksheet offenses has changed across the pre- and post-presumptive time periods. Among white offenders, Figure 2.7a shows that between Fiscal Years 2007 and 2008, the dispersion of sentence length shrank, particularly in the right tail of the distribution. The dispersion also narrowed in Fiscal Year 2013, prior to the shift to presumptive sentencing guidelines. In the following (post-presumptive) fiscal year, there appears to be a more dramatic shift in the 75^{th} percentile - sentence length fell from 84 months in Fiscal Year 2013 to 61 months in Fiscal Year 2015. At the 90^{th} percentile, sentence length for white offenders fell from 180 months to 120 months.

Among black property worksheet offenders, Figure 2.7b illustrates a slightly different pattern of change. The dispersion of total sentence length began to shrink in Fiscal Year 2010; sentence length at the 75^{th} percentile fell from 115 months to 84 months. And while the distribution of sentence length is tighter and shifted more to the left in Fiscal Year 2015 relative to Fiscal Year 2007, the distribution appears to be unaffected by the transition to presumptive

[54]Note, Fiscal Year 2007 does not include October through December 2006 (see Section 2.4).

[55]Alabama's voluntary sentencing guidelines were implemented at the beginning of Fiscal Year 2007 (October 2006), and the guidelines became presumptive at the start of Fiscal Year 2014 (October 2013).

[56]For legibility purposes, I cap total sentence length at 400 months in the box plots. A total sentence length of at least 400 months is not the maximum sentence length in the sample, but it is above the 99^{th} percentile of sentence length.

guidelines.

Figure 2.8 shows how the distribution of sentence length among individuals admitted for drug worksheet offenses has changed over time. Among white offenders, Figure 2.8a shows that following Fiscal Year 2009, the right tail of the sentence length distribution widened. For example, sentence length at the 75^{th} percentile increased from 97 months to 120 months. Yet, once the sentencing guidelines for drug worksheet offenses became presumptive, the dispersion of sentence length shrank, such that sentence length at the 75^{th} percentile fell from 120 months to 84 months.

Figure 2.8b illustrates that changes in the distribution of sentence length among black drug worksheet offenders followed a similar pattern to that of white offenders. Sentence length is relatively more dispersed in Fiscal Years 2010 to 2012. However, unlike the distribution among white offenders, the sentence length distribution among black offenders narrowed in Fiscal Year 2013 (in the pre-presumptive period) and continued to shrink such that sentence length at the 75^{th} percentile fell from 100 months to 60 months between 2007 and 2015. And like black property worksheet offenders, the entire distribution of sentence length shifted to the left.

Overall, Figures 2.7 and 2.8 illustrate that between 2007 and 2015, the most dramatic changes in the distribution of sentence length decisions occurred in the right tail. However, it is not always apparent that the shift to presumptive sentencing guidelines is the primary reason for the changing distributions, particularly among black worksheet offenders.

2.7.1 Methodology

To estimate the effects of race and the shift to presumptive sentencing guidelines across the distribution of sentence length, I estimate the following equation:

$$\begin{aligned} Y_{ict} = &\beta_1^\tau Post_t + \beta_2^\tau Black_{it} + \beta_3^\tau * t \\ &+ \beta_4^\tau Post_t^\tau * Black_{it}^\tau + \beta_5^\tau Black_{it} * t \\ &+ X_{it}'\beta_6^\tau + \gamma_c^\tau + \gamma_m^\tau + \gamma_o^\tau + \varepsilon_{ict}^\tau \end{aligned} \tag{2.6}$$

Let τ correspond to the $10^{th}, 25^{th}, 50^{th}, 75^{th}$, and 90^{th} percentiles. Let Y_{ict} represent the log total sentence length assigned to individual i, in county c, in quarter t. $Post_t^\tau$ is an indicator variable denoting the post-presumptive period. $Black_{it}^\tau$ is an indicator for whether an offender is black. Offender- and offense-level covariates, X_{it}^τ, include the number of counts for the first offense and indicators for whether the offender was convicted of a second or third offense, controlling for the offense type of the additional convictions. County fixed effects, γ_c^τ, month fixed effects, γ_m^τ, offense fixed effects γ_o^τ, and separate linear time trends (at the quarterly level) for white and black offenders are included. Standard errors are bootstrapped.[57]

2.7.2 Results

Tables 2.13 and 2.14 list the coefficients from the quantile regression estimated using the sample of property and drug worksheet offenders, respectively. The regression specification in Tables 2.13 and 2.14 is identical to the regression specification in Column 4 of Tables 2.10 and 2.11 (the OLS regressions on log total sentence length).

Recall, that in the regressions estimating average log total sentence length among property worksheet offenders (Table 2.10), black offenders received sentences that are statistically different from white offenders. However, in Table 2.13, at each percentile, the coefficient on $Black_{it}^\tau$ is not

[57] Standard errors are estimated using 100 bootstrap repetitions. The seed is 1001.

statistically significant. Only at the 75^{th} percentile is the coefficient on the time trend for black offenders statistically significant. Since the coefficient on the linear time trend for white offenders is also statistically significant, all else equal, each quarter sentence length at the 75^{th} percentile falls an additional 0.53 percent (or less than 1 month) for white offenders and an additional 1.51 percent (or 1.45 months) for black offenders.[58]

Furthermore, the coefficient on $Post_t^\tau$ is statistically significant at all but the 10^{th} percentile. The coefficient on $Post_t^\tau * Black_{it}^\tau$ is not statistically significant. In Table 2.13, the largest percentage change in sentence length occurs at the 90^{th} percentile; all else equal, sentence length falls by over 23 percent (or about 42 months) following the shift to presumptive sentencing guidelines. At the median, sentence length falls by 10.5 percent once the guidelines became presumptive. Note, only the coefficients on $Post_t^\tau$ at the 90^{th} and 50^{th} percentiles are statistically different from one another. In terms of magnitude, the OLS coefficient on $Post_t$ in Column 4 of Table 2.10 is similar to the coefficients at 25^{th} and 75^{th} percentiles. But it is smaller relative to the coefficient on $Post_t^\tau$ at the 90^{th} percentile and larger relative to the coefficient at the median.

Table 2.14 presents results from the quantile regressions using the sample of drug worksheet offenders. Only at the 90^{th} percentile is the coefficient on $Black_{it}^\tau$ statistically significant; all else equal, black offenders at the 90^{th} percentile in the pre-presumptive period have sentence lengths that are 11 percent larger (or about 20 months) relative to those of white offenders. At the 50^{th} and 75^{th} percentiles, the coefficient on the time trend for black offenders is statistically significant. For example, among black offenders, all else equal, sentence length at the 75^{th} percentile falls by an additional 0.8 percent each quarter.

Additionally, in Table 2.14, the coefficient on $Post_t^\tau$ is statistically significant at all but the 10^{th} percentile. Among white drug worksheet offenders, the largest percentage change in sentence length occurs at the 75^{th} percentile. All else equal, following the shift to presumptive sentencing guidelines, sentence length for white offenders at the 75^{th} percentile fell by over

[58]Sentence length for property worksheet offenders at the 75^{th} percentile is 96 months.

16.8 percent (or about 16 months). Furthermore, the coefficient on $Post_t^{\tau} * Black_{it}^{\tau}$ is statistically significant at the 90^{th} percentile. Among black offenders, all else equal, sentence length fell the most at the 90^{th} percentile once the guidelines became presumptive - it fell over 25.6 percent (or 46 months). Among white offenders, total sentence length at the 90^{th} percentile fell by about 11 percent. Because race may be correlated with legal factors that I cannot control for, the results are not necessarily indicative of changes in sentence length due to extralegal factors.

Lastly, among drug worksheet offenders, the OLS and quantile regression coefficients on $Post_t^{\tau}$ are relatively similar; in Column 4 of Table 2.11, the OLS coefficient on $Post_t$ is -0.143. The largest difference in the OLS and quantile regression coefficients occurs at the the 25^{th} percentile. Yet, the OLS regression does not capture how the change in guidelines differentially affects black offenders' sentence length, relative to that of white offenders.

In all, among property and drug worksheet offenders, the shift to presumptive sentencing guidelines is associated with a decrease in sentence length at all percentiles except the 10^{th}. And as expected, some of the largest changes in total sentence length occur in the upper tail of the sentence length distribution. Most notably, among drug worksheet offenders, the transition yielded different effects across race at the 90^{th} percentile. All else equal, the percent decline in black offenders' sentence length at the 90^{th} percentile was larger relative to the decline in white offenders' sentence length.

2.8 Conclusion

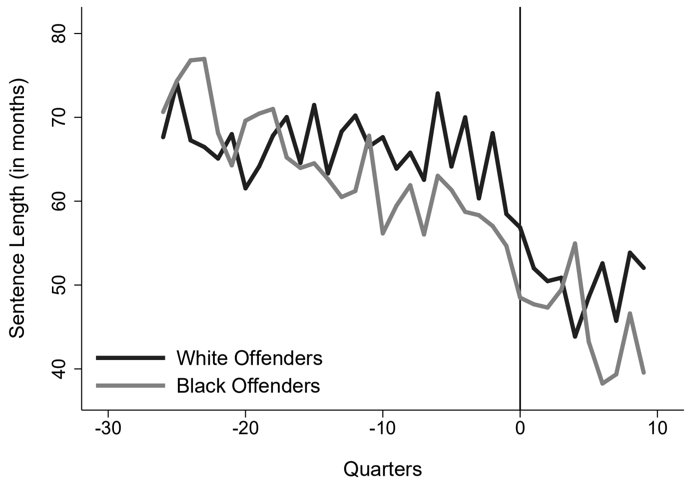
In October 2006, Alabama instituted voluntary sentencing guidelines that were expected to reduce overcrowding in prisons by sending fewer non-violent offenders to prison and reducing sentence lengths for non-violent offenders sentenced to prison. In October 2013, the guidelines became presumptive for property and drug worksheet offenses.

Since the guidelines aim to limit judicial discretion, which may change prosecutors' behavior; and they intend to reduce the use of incarceration among non-violent offenders, I estimate how, conditional on incarceration, the shift to presumptive sentencing guidelines may have affected offense and offender composition. I evaluate changes in: (1) admission type (new court commitment or parole/probation violation), (2) offense composition (such as average offense severity or the probability an offender enters prison with a second or third offense), and (3) inmate composition, namely age at admission and race. I find that the changes in composition are not consistent across offense type and race. Following the shift to presumptive sentencing guidelines, there is no statistically significant difference in the average number of property offenders admitted into prison, but all else equal, the average number of white drug offenders declines. Furthermore, among white drug offenders, there is a statistically significant decline in the average severity of admitted offenses. And while I find statistically significant increases in the average and median age at admission following the transition to presumptive guidelines, there are no statistically significant changes in racial composition.

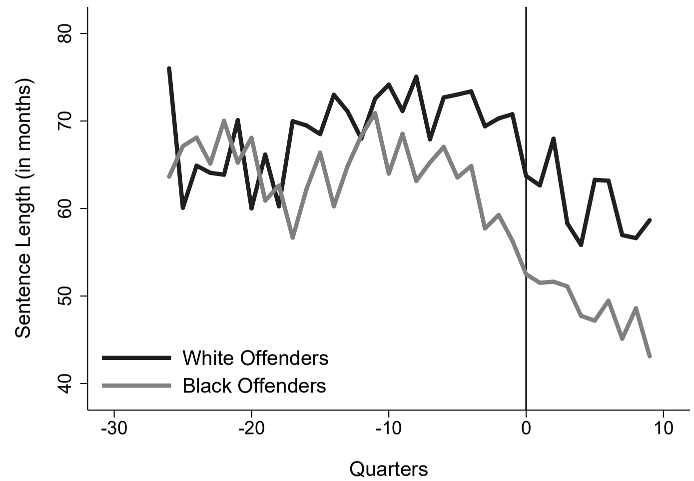
Next, I estimate a difference-in-differences model with personal worksheet admissions as the control group and find that after the guidelines became presumptive, average sentence length for property offenders fell 23 percent. Accounting for a difference in time trends across race, the transition does not have a statistically significant differential effect on the change in sentence length across white and black offenders. Relative to white personal worksheet offenders, sentence length among white drug worksheet offenders falls 11 percent. Among black offenders, there is no statistically significant difference in the percent change in sentence length between personal and drug worksheet offenders.

Lastly, I use quantile regressions to estimate how the effects of the transition varied across the distribution of sentence length. I find that among property and drug worksheet offenders, the shift to presumptive sentencing guidelines is associated with a decrease in sentence length at all percentiles except the 10^{th}. And as expected, total sentence length declines relatively more in the upper tail of the sentence length distribution. Additionally, among drug offenders, all else equal, the percent decline in black offenders' sentence length at the 90^{th} percentile was larger relative to white offenders' change in sentence length.

Given the limitations in using prison admissions data to evaluate the effects of the transition from voluntary to presumptive sentencing guidelines, future research could study how the guidelines affected the probability of incarceration and prosecutorial discretion. With more detailed information on previous criminal history, researchers could also evaluate the impact of extralegal factors in sentencing decisions as judicial discretion became increasingly limited.



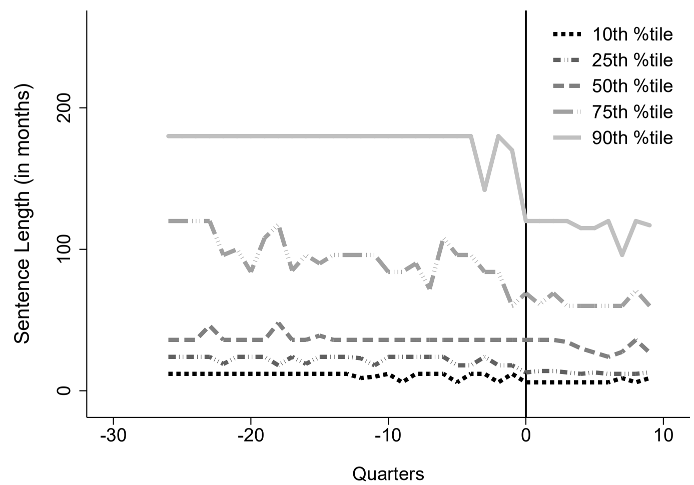
(a) Property WS Offenses



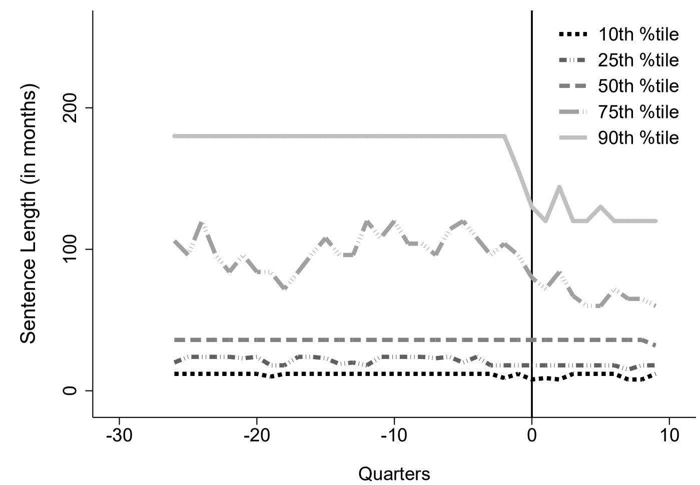
(b) Drug WS Offenses

Figure 2.1: Average total sentence length by race

This figure presents the average total sentence length at each quarter between January 2007 and December 2015. The graphs are generated separately for property and drug worksheet offenses. All worksheets were voluntary as of October 2006. In October 2013, property and drug worksheets became presumptive. Quarter 0 corresponds to the quarter before the property and drug offense worksheets became presumptive (July-September 2013). The sample is white and black males, aged at least 17 years old at the time of admission. The black solid line represents average sentence length among white offenders, and the gray solid line represents average sentence length among black offenders.



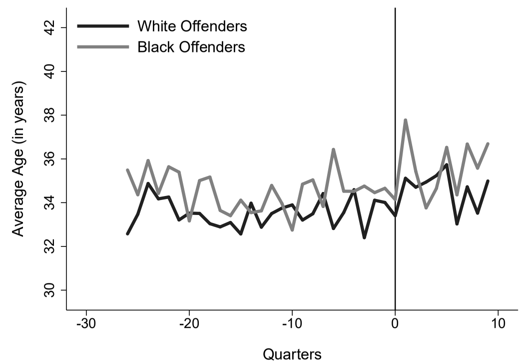
(a) Property WS Offenses



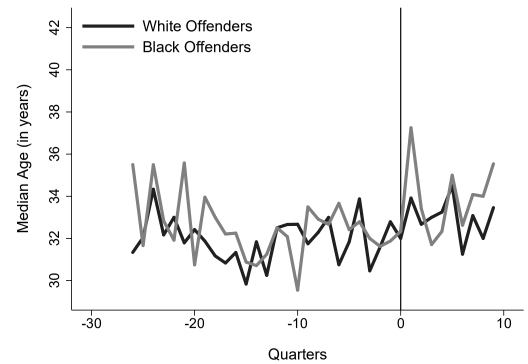
(b) Drug WS Offenses

Figure 2.2: Total sentence length by percentile

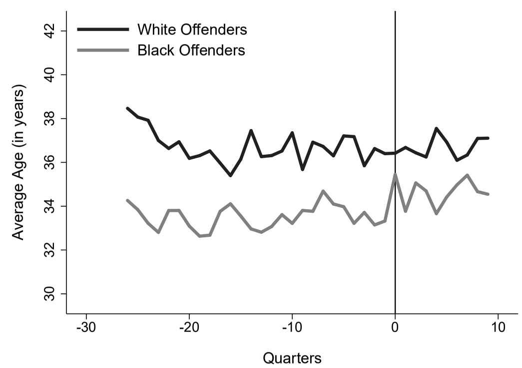
This figure presents the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles of total sentence length at each quarter between January 2007 and September 2016 by worksheet offense. The graphs are generated separately for property and drug worksheet offenses. All worksheets were voluntary as of October 2006. In October 2013, property and drug worksheets became presumptive. Quarter 0 corresponds to the quarter before the property and drug offense worksheets became presumptive (July-September 2013). The sample is white and black males, aged at least 17 years old at the time of admission.



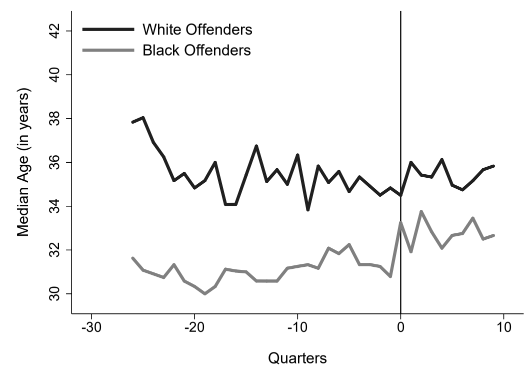
(a) Property WS



(b) Property WS



(c) Drug WS



(d) Drug WS

Figure 2.3: Average and median age at admission by race and worksheet type

This figure graphs average and median age by race and worksheet type. The black solid line corresponds to white offenders, and the gray solid line corresponds to white offenders. Quarter 0 corresponds to the quarter before the property and drug offense worksheets became presumptive (July-September 2013.)

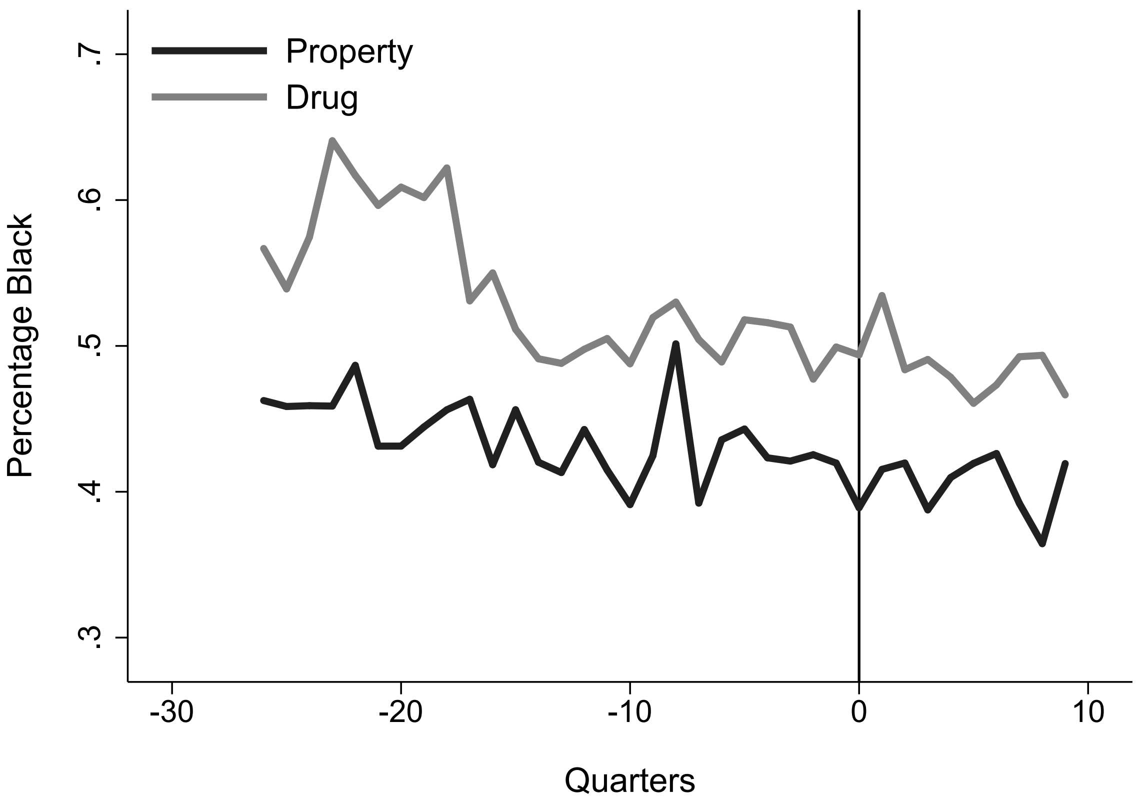
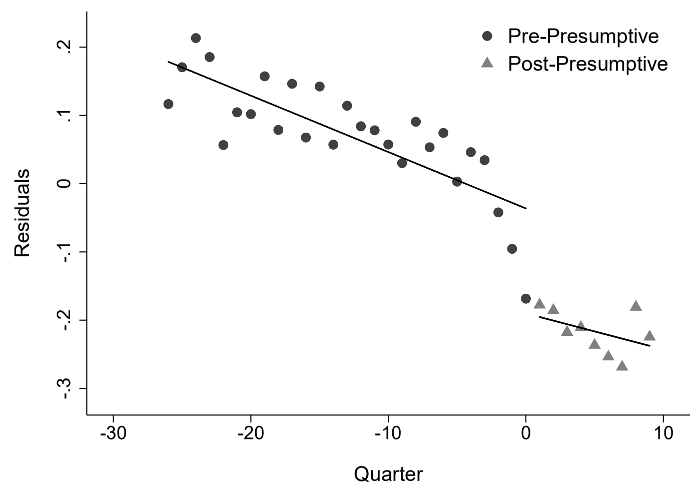
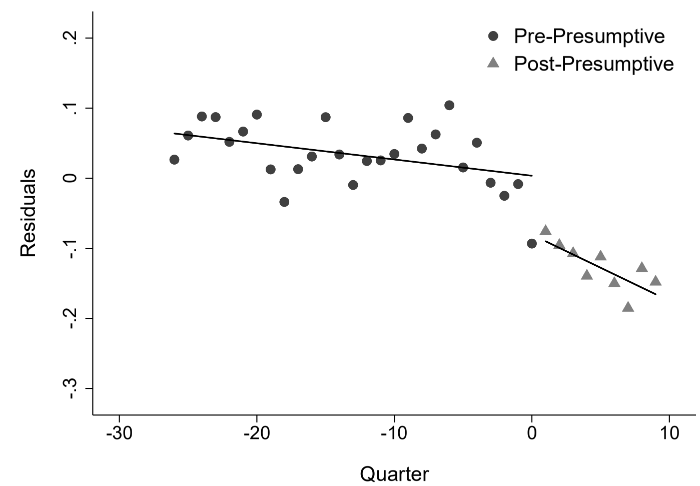


Figure 2.4: Share of black Offenders by worksheet type

This figure graphs the share of black offenders by worksheet type. The black solid line represents the share of property worksheet offenders that are black. The gray solid line represents the share of drug worksheet offenders that are black. Quarter 0 corresponds to the quarter before the property and drug offense worksheets became presumptive (July-September 2013).



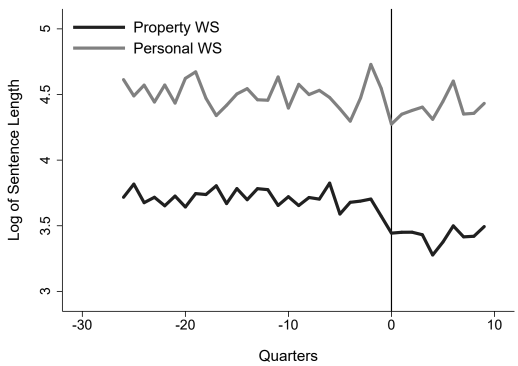
(a) Property WS



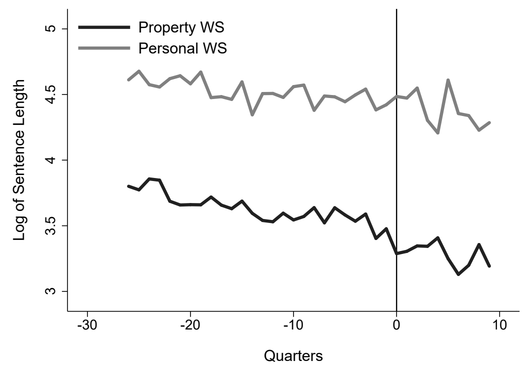
(b) Drug WS

Figure 2.5: Residuals from regression on log total sentence length

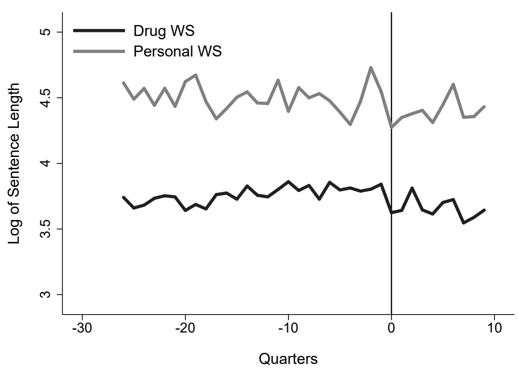
This figure presents residuals, averaged within a quarter, from a regression on log total sentence length. Regressions are run separately for property and drug worksheet offenders, and they control for race and county, month, and offense fixed effects. Admissions occurring between May and October 2013 are included. The black circle markers represent average residuals in the pre-presumptive period; the gray triangle markers represent average residuals in the post-presumptive period. A linear fit line, separately for the pre- and post-presumptive periods, are plotted. Quarter 0 corresponds to the quarter before the property and drug offense worksheets became presumptive (July-September 2013).



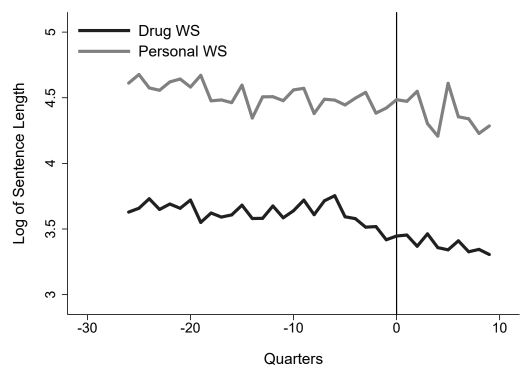
(a) Property WS, White Offenders



(b) Property WS, Black Offenders



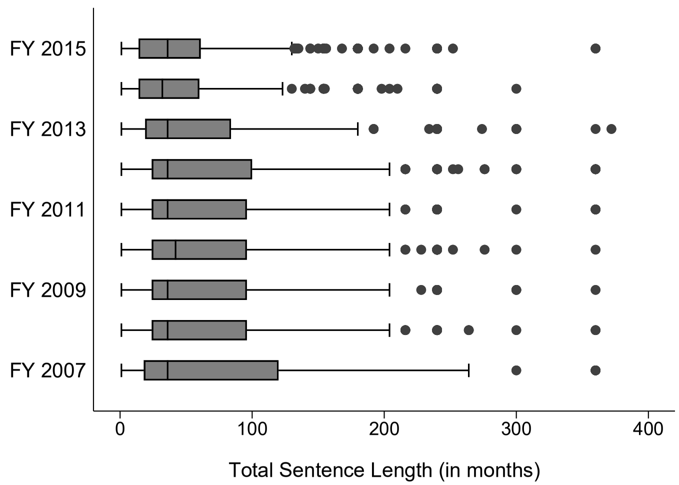
(c) Drug WS, White Offenders



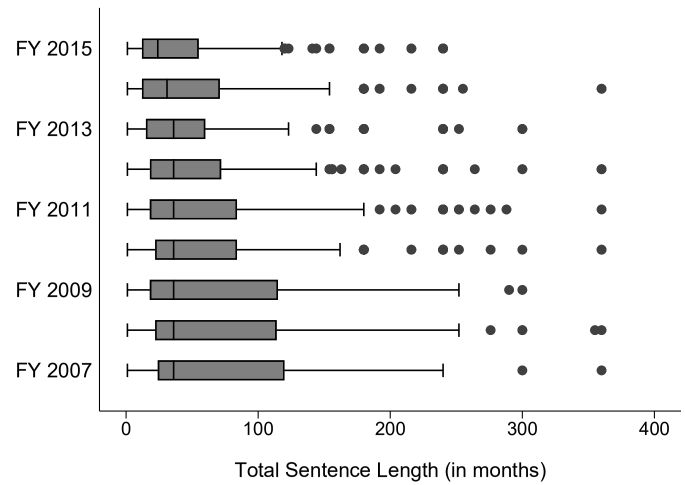
(d) Drug WS, Black Offenders

Figure 2.6: Graph of parallel trends

The top row of this figure compares a time-series of average log total sentence length among personal worksheet offenders and property worksheet offenders. Graphs in the bottom row compare a time-series of average log total sentence length among personal worksheet offenders and drug worksheet offenders. Graphs are generated separately by race. The black solid line represents average log total sentence length among property or drug worksheet offenders, and the gray solid line represents average log total sentence length among personal worksheet offenders. Quarter 0 corresponds to the quarter before the property and drug offense worksheets became presumptive (July-September 2013).



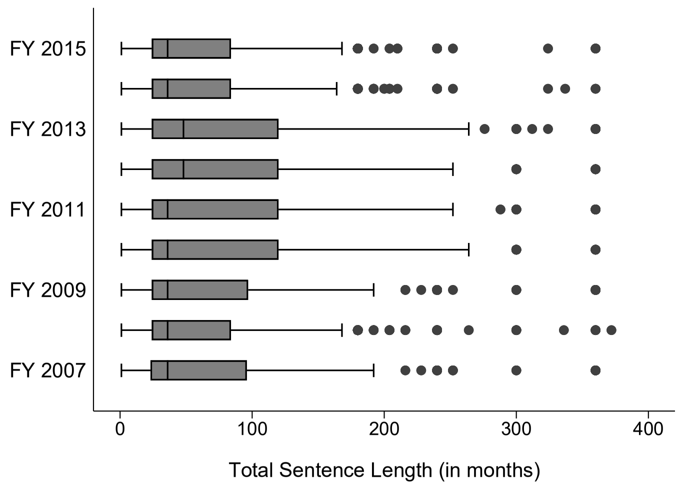
(a) White Offenders



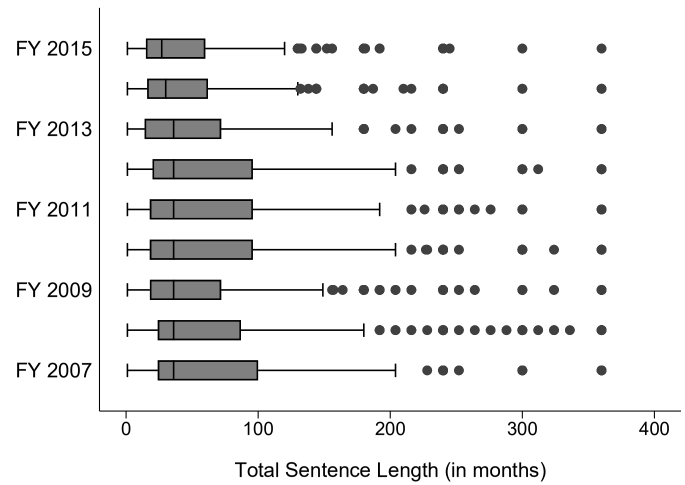
(b) Black Offenders

Figure 2.7: Distribution of total sentence length by fiscal year (FY) and race: property worksheet offenses

For each fiscal year between 2007 and 2015, this figure presents box plots of total sentence length by race among property worksheet offenders. Note, Alabama's voluntary sentencing guidelines were implemented at the beginning of Fiscal Year 2007 (October 2006), and the guidelines became presumptive at the start of Fiscal Year 2014 (October 2013). The length of each gray box represents the inter-quartile range, and the vertical line in the gray box denotes median sentence length. The whiskers represent 1.5 times the inter-quartile range, and the circle markers represent potential outliers. Total sentence length is capped at 400 months, and Fiscal Year 2007 does not include October through December 2006 (see Section 2.4).



(a) White Offenders



(b) Black Offenders

Figure 2.8: Distribution of total sentence length by fiscal year (FY) and race: drug worksheet offenses

For each fiscal year between 2007 and 2015, this figure presents box plots of total sentence length by race among drug worksheet offenders. Note, Alabama's voluntary sentencing guidelines were implemented at the beginning of Fiscal Year 2007 (October 2006), and the guidelines became presumptive at the start of Fiscal Year 2014 (October 2013). The length of each gray box represents the inter-quartile range, and the vertical line in the gray box denotes median sentence length. The whiskers represent 1.5 times the inter-quartile range, and the circle markers represent potential outliers. Total sentence length is capped at 400 months, and Fiscal Year 2007 does not include October through December 2006 (see Section 2.4).

Table 2.1: Descriptive statistics - all offenders

	Pre-Presumptive (1)	Post-Presumptive (2)
Property Offenses	0.20	0.22
	(0.398)	(0.416)
Drug Offenses	0.37	0.33
	(0.482)	(0.470)
Burglary Offenses	0.11	0.11
	(0.308)	(0.319)
Personal Offenses	0.21	0.20
	(0.408)	(0.401)
Other Offenses	0.12	0.13
	(0.324)	(0.338)
Property WS Offenses[a]	0.89	0.87
	(0.318)	(0.339)
Drug WS Offenses[b]	0.94	0.95
	(0.243)	(0.219)
Personal WS Offenses[c]	0.81	0.76
	(0.393)	(0.424)
N	51388	15280

The sample consists of white and black male offenders, aged at least 17 at the time of admission. All admissions are new court commitments. Descriptive statistics are calculated for two periods: (1) the pre-period, January 2007 to April 2013, and (2) the post-period, November 2013 to December 2015. Offenses are categorized as other, property, drug, burglary, or violent. For the first four rows, each offense category includes worksheet and non-worksheet offenses. "Other" offenses are unaffected by the voluntary and presumptive guidelines.

[a] The proportion of property worksheet offenses relative to all property offenses (not relative to all offenses regardless of type).

[b] The proportion of drug worksheet offenses relative to all drug offenses.

[c] The proportion of personal worksheet offenses relative to all personal offenses.

Table 2.2: Descriptive statistics - worksheet offenders

	Property WS Offense		Drug WS Offense		Personal WS Offense	
	Pre-Presumptive (1)	Post-Presumptive (2)	Pre-Presumptive (3)	Post-Presumptive (4)	Pre-Presumptive (5)	Post-Presumptive (6)
Total Sentence Length	65.73	47.97	66.68	54.37	165.82	152.28
	(65.230)	(50.018)	(71.620)	(53.731)	(236.510)	(212.358)
Black	0.44	0.40	0.54	0.48	0.69	0.68
	(0.496)	(0.491)	(0.498)	(0.500)	(0.464)	(0.466)
Ages 17-34	0.56	0.53	0.54	0.50	0.71	0.70
	(0.496)	(0.499)	(0.498)	(0.500)	(0.453)	(0.458)
Ages 35-49	0.36	0.37	0.37	0.41	0.22	0.22
	(0.480)	(0.483)	(0.483)	(0.492)	(0.416)	(0.417)
Ages 50+	0.08	0.10	0.09	0.09	0.07	0.08
	(0.264)	(0.305)	(0.284)	(0.291)	(0.247)	(0.264)
Age at Admission	33.95	35.05	34.97	35.70	30.15	30.74
	(9.947)	(10.197)	(9.837)	(9.619)	(10.914)	(10.682)
Offense Severity	47.37	47.93	72.99	75.14	292.57	282.58
	(8.487)	(8.521)	(17.565)	(17.881)	(201.331)	(199.985)
2^{nd} Offense	0.30	0.28	0.24	0.24	0.25	0.23
	(0.459)	(0.448)	(0.424)	(0.425)	(0.435)	(0.422)
2^{nd} & 3^{rd} Offense	0.20	0.19	0.08	0.09	0.11	0.12
	(0.402)	(0.393)	(0.266)	(0.285)	(0.315)	(0.323)
Court Commitment[a]	0.79	0.81	0.82	0.80	0.74	0.73
	(0.410)	(0.389)	(0.384)	(0.396)	(0.440)	(0.443)
N	8994	2942	17623	4794	8791	2354

The sample consists of white and black male offenders, aged at least 17 at the time of admission. All admissions are new court commitments. The pre-presumptive period is January 2007 to April 2013, and the post-presumptive period is November 2013 to December 2015. Columns 1 and 2 correspond to property worksheet offenders. Columns 3 and 4 correspond to drug worksheet offenders, and Columns 5 and 6 correspond to personal worksheet offenders.

[a] The share of individuals admitted for the commission of a new offense, relative to all admission types, such as parole or probation violations.

Table 2.3: Average total sentence length by worksheet and race

		Pre-Presumptive	Post-Presumptive
Property WS Offenses	White Offenders	66.53	50.10
		(64.540)	(50.859)
	Black Offenders	64.71	44.84
		(66.095)	(48.606)
Drug WS Offenses	White Offenders	69.29	60.03
		(70.077)	(55.859)
	Black Offenders	64.45	48.36
		(72.839)	(50.698)
Personal WS Offenses	White Offenders	170.03	168.12
		(262.622)	(277.944)
	Black Offenders	163.90	144.92
		(223.534)	(173.283)

Within each worksheet offense, this table separately lists the average total sentence length for white and black offenders. The sample consists of white and black male offenders, aged at least 17 at the time of admission. All admissions are new court commitments. Descriptive statistics are calculated for two periods: (1) the pre-period, January 2007 to April 2013, and (2) the post-period, November 2013 to December 2015. All offenses covered by worksheets were voluntary as of October 2006. In October 2013, property and drug worksheets became presumptive.

Table 2.4: Coefficients from regressions on admission type

	Property WS				Drug WS			
	New Commit		Violation		New Commit		Violation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.117	0.022	0.099	-0.003	-0.614**	-0.831**	0.487**	0.376
	(0.204)	(0.228)	(0.193)	(0.242)	(0.254)	(0.383)	(0.229)	(0.228)
Linear Time Trend	0.007	0.002	-0.008	-0.005	-0.003	0.004	-0.007	-0.003
	(0.011)	(0.010)	(0.006)	(0.006)	(0.016)	(0.017)	(0.007)	(0.006)
Black	-0.666	-0.488	0.198	0.066	0.506	0.263	0.654*	0.535*
	(0.525)	(0.613)	(0.257)	(0.297)	(0.983)	(0.998)	(0.368)	(0.319)
Black*Time Trend	-0.021*	-0.011	-0.027*	-0.035	-0.076***	-0.090***	-0.049***	-0.056***
	(0.012)	(0.013)	(0.016)	(0.022)	(0.021)	(0.026)	(0.015)	(0.020)
Black*Post		-0.330		0.244		0.451		0.225
		(0.326)		(0.245)		(0.462)		(0.246)
Joint F-test		0.306		0.242		0.197		0.042
Quarter FE	X	X	X	X	X	X	X	X
County FE	X	X	X	X	X	X	X	X
N	2615	2615	1327	1327	3268	3268	1874	1874

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from OLS regressions on the number of individuals admitted into prison either through a new court commitment (columns labeled "New Commit") or a parole/probation violation (columns labeled "Violation"). See Equation 2.1. Observations are at the quarter x county x race level. Regressions are run separately for property and drug worksheet offenses. All regressions include quarter and county fixed effects. Standard errors, clustered at the county level, are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t$ and $Post_t * Black_{it}$.

Table 2.5: Coefficients from regressions on the probability of admission type

	Property WS (1)	Drug WS (2)
Post	-0.010	-0.044**
	(0.017)	(0.019)
Linear Time Trend	0.001*	0.000
	(0.001)	(0.001)
Black	-0.013	-0.036**
	(0.013)	(0.016)
Black*Post	-0.035*	0.026
	(0.018)	(0.023)
Black*Time Trend	0.003***	0.001
	(0.001)	(0.001)
Joint F-test	0.005	0.302
Offense FE	X	X
Month FE	X	X
County FE	X	X
N	14829	27067

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from regressions estimating the probability of admission through a new court commitment relative to admission through a parole/probation violation (see Equation 2.2). Regressions are run separately for property and drug worksheet offenses. All regressions control for the number of counts for the first offense, whether the offender was convicted of a second or third offense, and month, county, and offense fixed effects. Standard errors, clustered at the county level, are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t$ and $Post_t * Black_{it}$.

Table 2.6: Coefficients from regressions on offense composition

	Pr(WS offense)		Severity		Pr(additional offense)	
	Property WS (1)	Drug WS (2)	Property WS (3)	Drug WS (4)	Property WS (5)	Drug WS (6)
Post	0.004	-0.031**	-0.143	-7.153***	-0.045**	0.006
	(0.008)	(0.014)	(0.364)	(0.945)	(0.019)	(0.016)
Linear Time Trend	0.001	0.001*	0.030	0.558***	-0.000	-0.000
	(0.000)	(0.000)	(0.019)	(0.045)	(0.001)	(0.001)
Black	-0.080***	0.004	-1.754***	-15.996***	-0.056**	0.003
	(0.012)	(0.015)	(0.424)	(1.138)	(0.026)	(0.022)
Black*Post	-0.007	0.031*	1.243*	7.284***	0.042	-0.009
	(0.009)	(0.018)	(0.638)	(1.066)	(0.033)	(0.021)
Black*Time Trend	0.000	-0.003***	-0.070**	-0.565***	-0.001	0.001
	(0.001)	(0.001)	(0.031)	(0.050)	(0.001)	(0.001)
Joint F-test	0.716	0.961	0.126	0.854	0.899	0.892
Offense FE					X	X
Month FE	X	X	X	X	X	X
County FE	X	X	X	X	X	X
N	66668	66668	11936	22417	11936	22417

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from three regressions. Columns 1 and 2 present coefficients from regressions on the probability an individual is admitted for a property or drug worksheet offense, relative to all offenses. Columns 3 and 4 present coefficients from regressions on average offense severity. Columns 5 and 6 present coefficients from regressions on the probability an individual is admitted with a second or third offense. Regressions are run separately for property and drug worksheet offenses. All regressions control for the number of counts for the first offense. Standard errors, clustered at the county level, are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t$ and $Post_t * Black_{it}$.

Table 2.7: Coefficients from regressions on age at admission - property worksheet offenses

	OLS Regression			Median Regression		
	All Offenders (1)	Black Offenders (2)	White Offenders (3)	All Offenders (4)	Black Offenders (5)	White Offenders (6)
Post	0.887**	1.240***	0.905***	1.378***	2.182***	1.382***
	(0.349)	(0.417)	(0.331)	(0.425)	(0.735)	(0.459)
Linear Time Trend	0.013	-0.016	0.010	-0.024	-0.036	-0.019
	(0.013)	(0.029)	(0.013)	(0.018)	(0.031)	(0.020)
Black	0.377			0.313		
	(0.635)			(0.484)		
Black*Post	0.323			0.390		
	(0.517)			(0.804)		
Black*Time Trend	-0.025			-0.005		
	(0.034)			(0.035)		
Joint F-test	0.010			0.010		
Offense FE	X	X	X	X	X	X
Month FE	X	X	X	X	X	X
County FE	X	X	X	X	X	X
Average/Median Age	33.95	34.53	33.5	32.08	32.38	31.83
N	11936	5138	6798	11936	5138	6798

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from regressions estimating the average and median ages at admission among property worksheet offenders. Columns 1 to 3 present coefficients from OLS regressions on age at admission. Columns 4 to 6 present coefficients from median regressions on age at admission. Columns 1 and 4 use the sample of all offenders and control for whether an offender is black. Columns 2, 3, 5, and 6 estimate the regressions separately by race. All regressions control for the number of counts for the first offense, whether the offender was convicted of a second or third offense, and month, county, and offense fixed effects. Standard errors, clustered at the county level, are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t$ and $Post_t * Black_{it}$.

Table 2.8: Coefficients from regressions on age at admission - drug worksheet offenses

	OLS Regression			Median Regression		
	All Offenders (1)	Black Offenders (2)	White Offenders (3)	All Offenders (4)	Black Offenders (5)	White Offenders (6)
Post	-0.032	0.431	0.108	0.185	0.692**	0.281
	(0.356)	(0.340)	(0.366)	(0.461)	(0.321)	(0.428)
Linear Time Trend	0.029**	0.057***	0.017	0.028	0.054***	0.014
	(0.014)	(0.016)	(0.013)	(0.021)	(0.013)	(0.020)
Black	-1.733***			-2.170***		
	(0.373)			(0.400)		
Black*Post	0.545			0.612		
	(0.396)			(0.563)		
Black*Time Trend	0.016			0.023		
	(0.019)			(0.025)		
Joint F-test	0.128			0.014		
Offense FE	X	X	X	X	X	X
Month FE	X	X	X	X	X	X
County FE	X	X	X	X	X	X
Average/Median Age	34.97	33.52	36.68	32.91	31.08	35.42
N	22417	11840	10577	22417	11840	10577

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from regressions estimating the average and median ages at admission among drug worksheet offenders. Columns 1 to 3 present coefficients from OLS regressions on age at admission. Columns 4 to 6 present coefficients from median regressions on age at admission. Columns 1 and 4 use the sample of all offenders and control for whether an offender is black. Columns 2, 3, 5, and 6 estimate the regressions separately by race. All regressions control for the number of counts for the first offense, whether the offender was convicted of a second or third offense, and month, county, and offense fixed effects. Standard errors, clustered at the county level, are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t$ and $Post_t * Black_{it}$.

Table 2.9: Coefficients from regressions on an indicator for whether an offender is black

	Property WS Offenses (1)	Drug WS Offenses (2)
Post	-0.0155	0.0081
	(0.0142)	(0.0116)
Linear Time Trend	0.0001	-0.0028***
	(0.0007)	(0.0004)
Offense FE	X	X
Month FE	X	X
County FE	X	X
N	11936	22417

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from regressions estimating the probability an offender is black, conditional on incarceration. Column 1 uses the sample of property worksheet offenders, and Column 2 uses the sample of drug worksheet offenders. All regressions control for the number of counts for the first offense, whether the offender was convicted of a second or third offense, and month, county, and offense fixed effects. Standard errors, clustered at the county level, are in parentheses.

Table 2.10: Coefficients from OLS regressions on log total sentence length - property worksheet offenses

	(1)	(2)	(3)	(4)	(5)
Post	-0.187***	-0.163***	-0.114*	-0.170**	-0.156***
	(0.053)	(0.059)	(0.062)	(0.075)	(0.052)
Linear Time Trend	-0.006***	-0.007***	-0.007***	-0.004	-0.004*
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Black	-0.063	0.054***	0.078***	-0.002	0.004
	(0.047)	(0.014)	(0.019)	(0.043)	(0.024)
Post*Time Trend			-0.002		-0.002
			(0.008)		(0.008)
Black*Post			-0.099*	0.011	
			(0.051)	(0.069)	
Black*Time Trend				-0.006*	-0.005**
				(0.003)	(0.002)
Joint F-test			0.000	0.007	
Offense FE	X	X	X	X	X
Month FE		X	X	X	X
County FE		X	X	X	X
N	11929	11929	11929	11929	11929

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from regressions on log total sentence length among property worksheet offenders. All regressions control for the number of counts for the first offense, indicators for whether the offender was convicted of a second or third offense (controlling for the offense type of the additional convictions), and offense fixed effects. Column 1 regresses log total sentence length without county and month fixed effects and no interaction terms. Column 2 introduces county and month fixed effects. Column 3 adds in $Post_t * Black_{it}$ and $Post_t * t$. Columns 4 and 5 include separate linear time trends for white and black offenders. Standard errors, clustered at the county level, are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t$ and $Post_t * Black_{it}$.

Table 2.11: Coefficients from OLS regressions on log total sentence length - drug worksheet offenses

	(1)	(2)	(3)	(4)	(5)
Post	-0.145***	-0.132***	-0.045	-0.143**	-0.087*
	(0.049)	(0.047)	(0.062)	(0.056)	(0.045)
Linear Time Trend	-0.003	-0.002	-0.002	0.000	0.001
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Black	-0.103	0.036	0.053**	-0.014	-0.009
	(0.066)	(0.024)	(0.022)	(0.050)	(0.040)
Post*Time Trend			-0.011		-0.011
			(0.007)		(0.007)
Black*Post			-0.080	0.010	
			(0.050)	(0.059)	
Black*Time Trend				-0.005	-0.004*
				(0.003)	(0.002)
Joint F-test			0.003	0.016	
Offense FE	X	X	X	X	X
Month FE		X	X	X	X
County FE		X	X	X	X
N	22398	22398	22398	22398	22398

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from regressions on log total sentence length among drug worksheet offenders. All regressions control for the number of counts for the first offense, indicators for whether the offender was convicted of a second or third offense (controlling for the offense type of the additional convictions), and offense fixed effects. Column 1 regresses log total sentence length without county and month fixed effects and no interaction terms. Column 2 introduces county and month fixed effects. Column 3 adds in $Post_t * Black_{it}$ and $Post_t * t$. Columns 4 and 5 include separate linear time trends for white and black offenders. Standard errors, clustered at the county level, are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t$ and $Post_t * Black_{it}$.

Table 2.12: Coefficients from difference-in-differences regressions on log total sentence length

	Property WS			Drug WS		
	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.094	0.040	0.044	0.044	-0.013	-0.013
	(0.063)	(0.042)	(0.069)	(0.059)	(0.047)	(0.058)
WS*Post	-0.237***	-0.233***	-0.233***	-0.122**	-0.115**	-0.118**
	(0.057)	(0.056)	(0.058)	(0.053)	(0.054)	(0.053)
Black	0.156***	0.087***	0.090**	0.166***	0.098***	0.096**
	(0.036)	(0.030)	(0.038)	(0.034)	(0.030)	(0.039)
WS*Black	-0.089**	-0.085**	-0.085**	-0.129***	-0.126***	-0.127***
	(0.038)	(0.033)	(0.036)	(0.043)	(0.043)	(0.043)
Black*Post	-0.095		-0.006	-0.094		-0.000
	(0.062)		(0.063)	(0.064)		(0.068)
WS*Black*Post	0.004		-0.001	0.009		0.006
	(0.062)		(0.061)	(0.091)		(0.091)
Linear Time Trend	-0.006***	-0.003*	-0.004	-0.003	-0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Black*Time Trend		-0.005*	-0.005		-0.005***	-0.005**
		(0.002)	(0.003)		(0.001)	(0.002)
Joint F-test	0.001		0.001	0.151		0.150
Offense FE	X	X	X	X	X	X
Month FE	X	X	X	X	X	X
County FE	X	X	X	X	X	X
N	23071	23071	23071	33540	33540	33540

${}^{*}p < 0.1$; ${}^{**}p < 0.05$; ${}^{***}p < 0.01$

This table lists the coefficients from difference-in-differences regressions on log total sentence length among property and drug worksheet offenders. In Columns 1 to 3, coefficients are estimated using the sample of property and personal worksheet offenders. In Columns 4 to 6, coefficients are estimated using the sample of drug and personal worksheet offenders. All regressions control for the number of counts for the first offense, indicators for whether the offender was convicted of a second or third offense (controlling for the offense type of the additional convictions), and month, county, and offense fixed effects. Columns 2, 3, 5, and 6 include separate linear time trends for white and black offenders. Standard errors, clustered at the county level, are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t * WS_{it}$ and $Post_t * Black_{it} * WS_{it}$.

Table 2.13: Quantile regression coefficients - property worksheet offenses

Percentile	10th	25th	50th	75th	90th
Post	-0.0953	-0.1751***	-0.1051**	-0.1636**	-0.2342***
	(0.0714)	(0.0548)	(0.0503)	(0.0644)	(0.0531)
Linear Time Trend	-0.0036	-0.0027	-0.0046**	-0.0053**	-0.0031
	(0.0030)	(0.0021)	(0.0021)	(0.0025)	(0.0023)
Black	0.0013	0.0573	0.0249	-0.0630	0.0088
	(0.0631)	(0.0468)	(0.0480)	(0.0583)	(0.0631)
Black*Post	-0.0000	-0.0875	-0.0504	0.0922	0.0336
	(0.1010)	(0.0784)	(0.0706)	(0.0872)	(0.0874)
Black*Time Trend	-0.0043	0.0005	-0.0043	-0.0098***	-0.0068
	(0.0043)	(0.0033)	(0.0031)	(0.0038)	(0.0042)
Joint F-test	0.259	0.000	0.008	0.302	0.004
Offense FE	X	X	X	X	X
Month FE	X	X	X	X	X
County FE	X	X	X	X	X
Pre-Presumptive % tile - Black Offenders	12	18	36	88	180
Pre-Presumptive % tile - White Offenders	12	24	36	120	180
N	11929	11929	11929	11929	11929

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from quantile regressions on log total sentence length using the sample of property worksheet offenders. All regressions control for the number of counts for the first offense, whether the offender was convicted of a second or third offense (controlling for the offense type of the additional convictions), and month, county, and offense fixed effects. A separate linear time trend for white and black offenders is also included. Bootstrapped standard errors are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t$ and $Post_t * Black_{it}$.

Table 2.14: Quantile regression coefficients - drug worksheet offenses

Percentile	10th	25th	50th	75th	90th
Post	-0.0828	-0.0838**	-0.1144***	-0.1684***	-0.1099***
	(0.0555)	(0.0399)	(0.0403)	(0.0367)	(0.0351)
Linear Time Trend	0.0011	0.0009	0.0012	-0.0007	-0.0033**
	(0.0019)	(0.0015)	(0.0018)	(0.0015)	(0.0017)
Black	-0.0141	0.0259	-0.0355	-0.0299	0.1113***
	(0.0417)	(0.0414)	(0.0437)	(0.0447)	(0.0393)
Black*Post	0.0228	-0.0564	0.0392	0.0530	-0.1461**
	(0.0736)	(0.0632)	(0.0601)	(0.0646)	(0.0576)
Black*Time Trend	-0.0011	-0.0008	-0.0064**	-0.0079***	0.0000
	(0.0025)	(0.0024)	(0.0026)	(0.0025)	(0.0024)
Joint F-test	0.174	0.001	0.085	0.010	0.000
Offense FE	X	X	X	X	X
Month FE	X	X	X	X	X
County FE	X	X	X	X	X
Pre-Presumptive % tile - Black Offenders	12	18	36	88	180
Pre-Presumptive % tile - White Offenders	12	24	36	120	180
N	22398	22398	22398	22398	22398

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists the coefficients from quantile regressions on log total sentence length using the sample of drug worksheet offenders. All regressions control for the number of counts for the first offense, whether the offender was convicted of a second or third offense (controlling for the offense type of the additional convictions), and month, county, and offense fixed effects. A separate linear time trend for white and black offenders is also included. Bootstrapped standard errors are in parentheses. The row labeled "Joint F-test" lists the p-value of an F-test testing the combined significance of $Post_t$ and $Post_t * Black_{it}$.

Chapter 3

No Such Thing as a (Correctional) Free Lunch? County-level Variation in Responses to Realignment and Proposition 47

3.1 Introduction

In 2011, the U.S. Supreme Court, in *Brown v Plata*[1], mandated that California reduce its prison population from 190 to 137.5 percent of design capacity within two years (a reduction of about 34,000 inmates). In response, California implemented a series of reforms. The first was SB 678 (July 2009), which incentivized counties to send fewer offenders to prison for probation failures. Because this only reduced the prison population to 179.5 percent of capacity, in October 2011, California implemented AB 109, referred to as public safety realignment. This reform shifted the supervision responsibility for low-level offenders from the state prison system to county

[1] 563 U.S. 493 (2011)

jails. Individuals convicted of "non-serious, non-violent, and non-sex-related" felony offenses were no longer sentenced to prison; rather, they served a short time in county jail or received non-custodial mandatory supervision (similar to probation). Within a year, the reform reduced the prison population by approximately 27,400 offenders; but the decline in the prison population still fell short of the court mandate. Therefore, in November 2014, California implemented Proposition 47, which re-classified low-level felonies as misdemeanors (which carry a maximum sentence of one year in county jail). The law also applied retroactively, such that individuals already serving prison or jail sentences for the affected offenses, at the time of passage, could petition to be released or have their sentences shortened.

Figure 3.1 illustrates a time-series of the prison, jail, and probation populations. Each vertical line represents the month in which a given reform was enacted. The black dashed line represents California's state prison population. The reforms were meant to reduce the prison population, and with each successive reform prison population is declining. The solid line represents the jail population. It decreases after SB 678[2], increases after Realignment, and declines after Proposition 47. The implementation of AB 109, or Realignment, was expected to increase the county jail population because individuals convicted of low-level offenses could no longer be sent to state prison; they had to be sanctioned locally. On the other hand, the jail population was expected to decline after Proposition 47 was adopted because re-classifying offenses as misdemeanors would reduce sentence lengths for low-level offenders. Lastly, the gray dashed line represents the number of individuals on probation. The number of people on probation fell after SB 678, levelled off following Realignment and fell following Proposition 47.[3]

[2]The jail population may be declining following SB 678 because parole and probation violators can also be sentenced to county jail. Therefore, if counties are reducing the rate at which they revoke probation or parole, then fewer offenders would be sent to county jail.

[3]Trends in probation populations are closely tied with trends in arrests. Prior to the passage of AB 109, total arrests had been declining. It is possible that probation populations did not continue to fall after Realignment because counties with little jail space available were required to increasingly use alternative sanctions. And arrests began to decline again after Proposition 47 was implemented.

Most papers study each of the reforms individually and address whether the reforms affected crime rates and recidivism. And while there are papers studying how AB 109 and Proposition 47 affect the offense- and offender-level composition of county jails, there is little information about the type of counties that were affected (in terms of county demographics and county budgets). Because jail capacity, preferences for traditional incarceration, and the costs of local supervision likely differ across counties, I study the variation in how counties were impacted by AB 109 and Proposition 47. I also examine whether Proposition 47 mitigated the effects of AB 109 in counties with the largest percent increase in the jail incarceration rate.

Figure 3.2 presents a scatter plot comparing the percent change in the prison and jail incarceration rates following Realignment.[4] There is a negative relationship between the two - the larger the percent decline in the prison incarceration rate, the larger the percent increase in the jail incarceration rate. In contrast to previous literature, I find no statistically significant relationship between the percent change in the prison and jail incarceration rates following AB 109. I also find that the fiscal impact (per resident) of AB 109 may have been largest for lower-income counties.

Figure 3.3 compares the percent change in each county's jail incarceration rate following AB 109 and the percent change in their jail incarceration rate following Proposition 47. There is a weak negative correlation suggesting that counties with larger percent increases in the jail incarceration rate following AB 109 also had larger percent declines in the jail incarceration rate following Proposition 47. Using a regression on the percent change in the jail incarceration rate following Proposition 47, I find that, consistent with Figure 3.3, changes in the jail incarceration rate following AB 109 are negatively associated with percent changes in the jail incarceration rate following Proposition 47, but the magnitude of the association differs across median income.

[4] I define the incarceration rate as the number of inmates per thousand county residents.

3.2 California's Policy Reforms

3.2.1 History of AB 109

In 2011, the U.S. Supreme Court ordered California to reduce its state prison population from 190 to 137.5 percent of design capacity. The decision was prompted by two court cases, *Coleman v. Brown*[5] and *Brown v. Plata*[6], which centered on inadequate medical and mental health care as a result of prison overcrowding. In order to comply with the order without having to release inmates or transfer inmates to county jails, the state legislature passed AB 109 (the Public Safety Realignment Act). Effective October 1, 2011, AB 109 shifted the supervision responsibility for most low-level offenders from the state to the county-level. It reallocated over 30,000 offenders, who would have otherwise been sentenced to state prison, to county jail or community supervision.[7]

AB 109, or Realignment, shifted the responsibility for three criminal justice populations from the state to the county-level: (1) newly convicted low-level offenders, (2) technical violators of probation or parole, and (3) inmates in state prison that would have been released to state parole. These populations consist of individuals with current and/or previous convictions of nonserious, nonviolent, and nonsexual offenses (the "non-non-non's"); they cannot have any prior convictions for serious, violent, or sexual offenses. Among newly convicted offenders, individuals whose most serious offense at conviction was one of more than 500 felonies could no longer be sentenced to state prison. Instead, they had to be sentenced to county jail or probation.[8] Furthermore, once an offender?s jail sentence was served, they were released without restrictions

[5]No. 2:90-cv-0520 LKK JFM (PC) (E.D. Cal. Aug. 19, 2011)

[6]563 U.S. 493 (2011)

[7]AB 109 was also designed to improve re-entry and recidivism outcomes since offenders would be closer to their families and community-based services. It was believed that local, rather than state, authorities know how to better address the needs of offenders and that inter-agency coordination between local sheriffs, probation departments, and social services agencies would ease the re-entry process (Petersilia 2014; Lofstrom and Raphael 2013a).

[8]AB 109 did not affect sentence length; offenders sentenced to county jail can receive the same sentence length they would have received had they been sentenced to state prison.

or supervision.[9]

Additionally, AB 109 mandated that most "non-non-non" offenders that violated the terms of their probation or parole could no longer be sentenced to state prison. Rather, as with newly convicted offenders, they must be sentenced to county jail or receive post-release community supervision (PRCS), which is similar to probation. AB 109 also reduced the maximum penalty received for violations: prior to AB 109, the maximum prison sentence received by technical violators was one year; whereas, following AB 109, the maximum penalty was 6 months in county jail.

Another provision of AB 109 shifted state parole responsibility for non-violent offenders to county probation departments. Prior to AB 109, most (low-level) offenders who completed their prison sentences were paroled to their home counties and supervised by state parole agents. Following AB 109, these offenders were now supervised by county probation departments. Furthermore, the length of supervision declined. Before Realignment, parole supervision lasted for up to 3 years (offenders could be discharged after 13 months if they had no new violations). After AB 109 was adopted, individuals on PRCS were eligible for discharge at 6 months (Petersilia and Snyder 2013).

AB 109 also provided counties with new alternatives to traditional incarceration and probation. Felony offenders could receive a split sentence, which is a combination of a term in county jail and mandatory supervision.[10] And certain offenders under probation could receive "flash incarceration,? in which an offender that violates community supervision is sentenced for up to 10 days in county jail without undergoing the process for traditional revocation (Petersilia and Snyder 2013).

Lastly, because the fiscal responsibility for a large number of offenders was suddenly shifted to counties, California dedicated $2.3 billion to be allocated to counties over the first three years of realignment. Initially, fund allocations were calculated using a formula that

[9]This is true only for offenders who did not receive a split sentence.

[10]Mandatory supervision follows the same terms, conditions, and procedures of traditional probation.

took into account the number of offenders estimated to be realigned to the counties (60%), a county?s adult population (30%), and how much a county reduced its prison incarceration rate, under SB 678, prior to AB 109 (10%) (Bird and Hayes 2013). However, counties with low pre-AB 109 prison incarceration rates were at a disadvantage ? because they would have smaller realignment population projections, they would receive less assistance. Since the first-year allocation formula only captured the shift in the fiscal responsibility from the state to counties at the time of realignment, rather than the total burden counties would incur managing a larger offender population over time, the allocation formula in the second year of realignment changed. Second- and third-year allocations were calculated by taking the maximum of four options: doubling the first year?s allocation, using the first-year allocation formula with updated population counts, using only the size of the projected realignment population, or using only the size of a county's adult population.

3.2.2 History of Proposition 47

However, AB 109 did not sufficiently reduce the prison population below the court-mandated target (Bird et al. 2018). Therefore, effective November 14, 2014, Proposition 47 (Prop 47) re-classified a set of low-level drug and property offenses from felonies to misdemeanors and sentenced misdemeanants up to one year in jail.[11,12] These new misdemeanor provisions did not apply to offenders who had one or more convictions for violent felonies.

Additionally, offenders that were incarcerated for the re-classified offenses before November 2014 could petition the original trial court for resentencing. If the petitioner was determined eligible and did not pose an unreasonable risk of danger to public safety, the court would recall the sentence and impose a new one under the new misdemeanor provisions. The resentenced

[11] Some of these offenses were considered wobblers, offenses that could be charged either as felonies or as misdemeanors (at prosecutors' discretion) .

[12] Proposition 47 re-classified the following felonies to misdemeanors: drug possession, shoplifting, receiving stolen property, forgery of checks, bonds or bills, writing checks with insufficient funds, and petty theft.

prison term imposed could not be longer than the original one, and resentenced petitioners were given credit for time served.[13] For individuals who already completed their sentences for Prop 47 offenses, Proposition 47 gave them the opportunity to reduce the felony conviction to a misdemeanor.

The California Legislative Analyst's Office (LAO), in Taylor (2015), estimated that each year 40,000 offenders are convicted of Prop 47 offenses. Between the adoption of AB 109 and the adoption of Prop 47, approximately 90 percent of offenders that received felony convictions for Prop 47 offenses were sentenced to county jail and/or community supervision. The remaining 10 percent received prison sentences. Therefore, the effect on counties' criminal justice workload was twofold. On the one hand, the 10 percent of offenders that once received prison sentences became the county's responsibility. On the other hand, Prop 47 was expected to free up bed space in county jails since offenders convicted of Prop 47 offenses serve shorter sentences in county jail (a maximum of one year relative to a sentence between 16 months and 3 years). Furthermore, individuals on trial for Prop 47 offenses were less likely to be held in jail as they awaited the conclusion of their trial (since the offenses were now misdemeanors). But while Prop 47 had the ability to reduce overcrowding (and mitigate the impacts of Realignment), if capacity-constrained jails used the additional bed space to reduce early releases, the effects of Prop 47 on the size of the jail population would have been smaller.

3.3 Literature Review

There has been considerable research, at the state-level, on California's criminal justice reforms - particularly AB 109. There are also some studies that examine county-level variation in responses to AB 109. They primarily examine how the jail population has changed, both in terms of offender and offense characteristics. I have not come across a paper that assesses

[13]Resentencing did not apply to individuals with one or more convictions of certain crimes, including but not limited to serious or violent felonies and/or sex offenses.

the county-specific fiscal impact of AB 109 or Proposition 47. Nor have I found research on how Proposition 47 mitigates the effects of AB 109 (at the county-level). Typically, research has focused on whether the reforms affected crime rates and recidivism, sometimes exploiting county-level variation in pre-reform prison incarceration rates.[14]

3.3.1 AB 109

AB 109, or Realignment, led to an immediate decline in prison admissions. Prison releases were slower to fall, but within a year admissions and releases had balanced out and, as in Figure 3.1, the prison population stabilized (Lofstrom and Raphael 2016). Studying county-level variation in jail capacity constraints and counties' response to AB 109, Lofstrom and Raphael (2013a), show that decreases in prison incarceration rates led to increases in jail incarceration rates, but not at a one-to-one rate. They find that, on average, a county?s jail population increased by one inmate for every three individuals no longer sentenced to state prison.

Lofstrom and Raphael (2013a) also provide evidence of jail crowd-out effects due to Realignment. They find that individuals convicted of felonies and parole violations were displacing pre-trial (or unsentenced) inmates and individuals serving time for misdemeanor offenses. There is also evidence of increases in capacity-constrained (or early) releases, particularly in counties with court-ordered population caps.[15] In counties with population caps, because of capacity constraints, for every four realigned offenders, one sentenced inmate per month was released early. In counties without a court-ordered population cap, one sentenced inmate for every 16

[14]Lofstrom and Raphael (2013b), Lofstrom and Raphael (2016), and Sundt, Salisbury, and Harmon (2016) find no evidence that the passage of AB 109 affected violent crime rates. There appears to be small increases in property crime rates, primarily among auto thefts. Bird and Grattet (2017) find that, following AB 109, statewide recidivism rates slightly increased. Following the passage of Proposition 47, Bartos and Kubrin (2018) find no impact on violent crimes, but it may have led to increases in larceny and auto thefts. Bird et al. (2018) find that recidivism fell after Proposition 47.

[15]There are 18 counties with a court-ordered cap on jail populations and 20 counties with self-imposed caps. Population caps are imposed at the jail level, not the county level. As of March 2014, of the 119 county jail facilities, 39 were under a court-ordered population cap. Population caps have been in place for decades and are rarely terminated (Lawrence 2014).

realigned offenders was released early (Lofstrom and Raphael 2013a).[16]

Verma (2016) also studies county-level variation in responses to AB 109. But rather than using a continuous measure of the change in the jail incarceration rate, Verma uses a binary measure of decarceration (a decrease in both the jail and prison incarceration rates). She divides counties into groups based on their pre-AB 109 trajectories of prison incarceration rates and finds that counties with a consistently low incarceration rate were more likely to decarcerate. Verma also finds that decarceration is the most likely response among counties with the highest prison incarceration rates. Moreover, in contrast to Lofstrom and Raphael (2013a) her results suggest that jail occupancy was not a significant factor in shaping pre-AB 109 prison incarceration rate trajectories, and it did not appear to predict decarceration.

In addition to studies of state- and county-level variation in offender and offense characteristics, Lin and Petersilia (2014) trace how counties used the funds they received as part of Realignment. They find that counties that allocated more of the funds to law enforcement were responding to local criminal justice needs, such as higher law enforcement expenditures, a higher incarceration rate for drug felonies, and higher serious crime rates. Counties that allocated relatively more funds to treatment programs and social services had higher rates of black unemployment and lower probation expenditures. The authors did not find that median income, county revenues per resident, operational expenditures for jails, or jail overcrowding had a statistically significant association with a county's decision on how to allocate AB 109 funds on enforcement and treatment programs.

Within the literature on AB 109, like Lofstrom and Raphael (2013a), my results suggest that jail capacity constraints play an important role in how counties respond to an increasing offender responsibility. And, in addition to increased use of early release, I provide evidence that capacity-constrained counties may also be using relatively more probation to manage their offender caseload. However, I do not find the change in prison incarceration rates to predict

[16]Capacity-constrained releases are not a result of Realignment, but their use increased after AB 109 was implemented (Turner, Fain, and Hunt 2015).

changes in the jail incarceration rate. I, instead, find that the pre-AB 109 share of offenders sent to prison, relative to jail, has a statistically significant association with the percent change in jail incarceration rate.[17] Additionally, this paper further contributes to the literature by providing an estimate of the fiscal impact of AB 109 on each county, and it assesses whether low-income counties were disproportionately affected by AB 109.

3.3.2 Proposition 47

Fewer papers study the county-level variation in responses to Prop 47, which re-classified a subset of drug and property offenses from felonies to misdemeanors. Like the research on AB 109, the literature on Prop 47 describes the effects of re-classification on offense and offender composition in county jails and describes the role that jail capacity constraints may play in the change in the jail population.

Nguyen et al. (2016) show that following the implementation of Prop 47, the average daily jail population fell by about 8,000 inmates, which brought the entire California jail system below its rated capacity for the first time since the passage of AB 109. They also find that the number of individuals serving sentences for Prop 47 offenses in county jail declined by 50 percent, which contributed to a 9 percent decline in the total jail population in the year following Prop 47. Furthermore, as suggested by Taylor (2015), there is evidence that capacity-constrained jails used some of the newly freed jail space to house offenders they would have otherwise had to release early (Nguyen et al. 2016; Grattet et al. 2017).

In addition to studies analyzing the changes in jail composition, some papers have projected the fiscal savings counties would accrue following the passage of Prop 47. Males and Buchen (2014) predict the savings for three large California counties by estimating the contribution of each county to the total number of people affected by Prop 47. They estimate

[17] Note, the dependent variable in Lofstrom and Raphael (2013a) is a change in (levels of) prison incarceration rates. I use the percent change in prison incarceration rates as the dependent variable.

annual statewide cost savings (as a result of freed jail space) to range between $400 to $700 million. Romano (2015) uses the average annual marginal cost of a jail inmate (estimated by the California Department of Corrections and Rehabilitation) and the estimated annual decline in the number of jail inmates to calculate that county jails could have expected total savings of $203 million. Unlike these papers, I use the measured changes in county jail populations to estimate the county-by-county fiscal impacts of Proposition 47. I also contribute to the literature on Proposition 47 by exploring how it might mitigate the effect of increasing jail incarceration rates (and, therefore, increasing fiscal responsibility) in counties most affected by Realignment.

3.4 Data

There are 58 counties in California. I omit Alpine, Sierra, and San Francisco counties from all analyses.[18] To study changes to California's criminal justice system, I combine data from a variety of sources. The total adult population in prison as of December 31 is provided by the California Sentencing Institute, a project of the Center on Juvenile and Criminal Justice (CJCJ).[19] Monthly data on the average daily (jail) population and the number of inmates released from jail is provided by the Board of State and Community Corrections' (BSCC) Jail Profile Survey. Monthly data on adult probation populations is provided by Open Justice, which collects data maintained by California's Department of Justice (DOJ).

Recall that AB 109 became effective in October 2011. Because state prison populations are only available for December, I keep comparisons consistent across prison, jail, and probation populations by using the December average daily population and the total probation caseload at

[18] Alpine does not maintain its own county jail system (Lofstrom and Raphael 2016), and the State Controller's Office (SCO) does not provide information for San Francisco that is comparable to the data reported for all other counties (California Budget & Policy Center 2017). I do not have complete criminal justice data for Sierra County.

[19] County prison population data refers to the county where an individual was sentenced. This is the same county where the crime was committed, and is usually the county an offender lives in. Upon release from state prison, inmates are returned to their county of commitment. Changes in prison incarceration rates at the county level are generally representative of changes in the extent to which counties are incarcerating their local residents in the state system. 90 percent of individuals released from prison return to the county where they were sentenced (Beck 2006).

the end of December. To estimate percent changes in the jail incarceration rate coinciding with Realignment, I use the difference in average daily jail populations (per county resident) as of December 2010 and December 2012.[20] Similarly, I estimate the percent change in the state prison population (per county resident) between December 2010 and December 2012. The percent change in the probation population is calculated in the same manner. To estimate the percent change in releases following AB 109, I compare the total number of releases (per county resident) in October 2010 through September 2011 to the total number of releases (per county resident) in October 2011 through September 2012.[21]

Prop 47 became effective in November 2014. To estimate percent changes in the jail population (per county resident) following its implementation, I use the difference in average daily populations (per county resident) as of December 2013 and December 2015. I estimate the percent change in the state prison population (per county resident) between December 2013 and December 2015. To estimate the percent change in releases following Prop 47, I compare the total number of releases (per county resident) in November 2013 through October 2014 to the total number of releases (per county resident) in November 2014 through October 2015.

Because some county jails faced capacity constraints prior to AB 109, they may not have been able to absorb offenders (that would have previously been sentenced to prison) without increasing jail capacity or releasing inmates early. Therefore, I include a measure of the jail population-to-capacity ratio. Using rated design capacities provided by the BSCC[22], I estimate county-level, pre-AB 109 jail population-to-capacity ratios by taking the ratio of the December 2010 average daily jail population and the rated capacity as of July 2010. I calculate the post-AB 109 (pre-Prop 47) jail population-to-capacity ratio by taking the ratio of the December 2013 average daily population and the July 2013 rated capacity.

To describe the counties affected by the passage of AB 109 and Prop 47, all county-level

[20] I choose December 2012, rather than December 2011, since it took about a year for AB 109 to achieve full effect in terms of prison and jail population changes.

[21] I define releases as the sum of pre-trial releases and capacity-constrained releases.

[22] Source: Rated Capacities of Type II, III, and IV Local Adult Detention Facilities

characteristics are pre-AB 109.[23] This includes median income, the share of the county population that is black or Hispanic, arrests per thousand residents, the percent of the county population represented by the urban population, the number of probationers per thousand residents, and the number of jail and prison inmates per thousand residents.[24] I also include a county-level index for racial heterogeneity using a measure similar to the Theil index. Heterogeneity is defined as the absolute value of $\sum_i p_i log(p_i)$, where p_i is racial group i's share of the population.[25] Values closer to one in absolute value indicate greater racial heterogeneity.

I also create pre-AB 109 measures of prison and probation use. I estimate the share of offenders counties housed in state prison, relative to the sum of prison and jail inmates. This measures the extent to which a county used state prison, rather than jail, as a way to incarcerate offenders. Additionally, I estimate the share of offenders receiving probation, relative to the sum of prison inmates, jail inmates, and probationers. This measures a county's use of alternative sanctions.

Furthermore, I estimate the impact AB 109 and Prop 47 may have had on county budgets. For example, the estimate of the fiscal impact of AB 109 is meant to provide a rough approximation of the amount jail expenditures would have had to increase to accommodate the rise in offender caseload. County budgets are reported by the State Controller's Office (SCO). To measure total county-level spending per resident, I use "total expenditures during the fiscal year" from Fiscal Year 2011 (July 2010 to June 2011).[26] I use county spending on adult detention centers as a measure of jail expenditures.[27]

[23] For annual-level characteristics, I use 2010 data. For values that are reported monthly, I use observations from December 2010.

[24] Sources: Median income, in 2010 dollars, is from the Small Area Income and Poverty Estimates (SAIPE). County population estimates are from the U.S. Census Bureau's Annual Estimates of the Resident Population by Sex, Race Alone or in Combination, and Hispanic Origin: April 1, 2010 to July 1, 2018. Arrest data is provided by Open Justice using data from the Monthly Arrest and Citation Register (MACR). The percent of the the total population of the county represented by the urban population is listed in the 2010 Census Urban Area List Files, 2010 Percent Urban and Rural by County.

[25] All classifications of race (listed in the intercensal estimates) are used to calculate the heterogeneity index.

[26] See California Budget & Policy Center (2017).

[27] See California Budget & Policy Center (2017). I only use the county's operational budget due to missing data on capital outlays.

To estimate the fiscal impact of the reforms on each county, I first calculate the average pre-AB 109 cost per jail inmate (or jail expenditures divided by the average daily jail population in 2010). Then, I estimate the fiscal impact of AB 109 by multiplying the average cost per jail inmate and the change in the number of jail inmates (between December 2010 and December 2012). Similarly, to calculate the fiscal impact of Prop 47, I multiply the average pre-AB 109 cost per jail inmate and the change in the number of jail inmates between December 2013 and December 2015.

Lastly, in addition to estimates of fiscal impact, I use data on the first-year allocation of AB 109 funds provided to each county. County-level data on AB 109 first-year allocations is listed in Bird and Hayes (2013). To assess the extent to which the AB 109 funds mitigated the fiscal impact of sudden jail population increases, I use the sum of the funds designated for managing the "realigned" offender population and for start-up costs associated with increasing jail capacity.

3.5 County Descriptive Statistics

Before discussing county-level variation in the effects of AB 109 and Prop 47, I group counties by median income and describe county-level demographics, county budgets, and counties' criminal justice systems. Table 3.1 stratifies counties into three equally-sized groups according to the distribution of median income. The table is divided into three panels: (1) county demographics, (2) county-level fiscal measures, and (3) county-level criminal justice indicators.

Panel A illustrates that, in California, counties in the bottom third of the distribution of median income are more rural and less racially diverse than counties in the middle and top thirds of the distribution. Panel B lists the counties' pre-AB 109 total expenditures and jail expenditures (averaged within terciles of median income). Moving from the bottom to the top third of median income, total county expenditures per resident declines. One possible explanation is that in counties with higher median incomes a larger share of the population is concentrated in urban areas. And in areas that are more densely populated, the city, rather than the county, covers the cost of municipal services.

Next, Panels B and C show that jail expenditures among the top third of counties, where median income is high, make up a larger share of total expenditures. However, on average, these counties have the lowest number of jail inmates per thousand residents. In addition, counties in the top third of the median income distribution have the largest jail expenditures per jail inmate (the average cost of a jail inmate), but the lowest jail expenditures per resident. This can not be the result of unused jail space (which would contribute to relatively higher overhead costs per inmate), because, on average, counties in the top third of median income have a higher pre-AB 109 jail population-to-capacity ratio. Instead, a possible explanation is that property values and labor costs are higher in counties with a higher median income, and these relatively more expensive jails in high median income counties are occupied by fewer inmates.

Lastly, to illustrate how median income is distributed across counties categorized by their incarceration rates, Figure 3.4 presents a scatter plot of counties' pre-AB 109 prison and jail incarceration rates. The vertical and horizontal lines represent the median prison incarceration rate and the median jail incarceration rate, respectively. The black circle-markers represent counties that fall below the median of median income, and the gray circle-markers represent counties that lie above the median. This graph illustrates, like Panels A and C of Table 3.1, that counties with larger median incomes use prison and jail less (possibly because of lower crime rates). Counties below the median of median income generally lie in the "above median prison and jail incarceration" rates quadrant. Recall that Lofstrom and Raphael (2013a) show that counties with higher pre-AB 109 prison incarceration rates experienced larger changes in prison incarceration rates following the adoption of AB 109. Therefore, combined with Figure 3.4, this suggests that, on average, counties with low median income had a larger number of offenders (per resident) to absorb into county supervision once AB 109 was implemented.

3.6 AB 109

Effective October 1, 2011, AB 109 shifted the supervision responsibility for most low-level offenders from the state to the county-level by narrowing the set of felony offenses that could receive a prison sentence. Figure 3.2 compares the percent change in each county's prison and jail incarceration rates following AB 109. It illustrates that counties that experienced larger percent declines in the prison incarceration rate also experienced larger percent increases in their jail incarceration rate. However, as discussed in Lofstrom and Raphael (2013a) and illustrated in Figure 3.2, the magnitude of the change in the jail incarceration rate is less than the magnitude of the change in the prison incarceration rate (this is true in both levels and percent change). This suggests that variation in counties' demographics, economies, and use of state prison (or incarceration in general) can contribute to variation in the extent to which county jails absorb realigned offenders.

3.6.1 AB 109 and the Criminal Justice System

I begin by describing counties' criminal justice features based on where they lie in the distribution of the percent change in the jail incarceration rate (or the number of jail inmates per thousand residents). Table 3.2 stratifies counties into three equally-sized groups according to the distribution of the percent change in jail incarceration rates following AB 109. The table is divided into two panels: (1) pre-AB 109 criminal justice characteristics and (2) changes in criminal justice populations (and behavior) following Realignment.

On average, counties in the bottom third of the distribution of changes in jail incarceration rates experienced declines in their jail incarceration rate. Counties in the top third experienced (percent) increases in their jail incarceration rates that were about 4 times larger than increases among counties in the middle third. Panel A shows that there is a positive association between the percent change in the jail incarceration rate and the percent of incapacitated offenders that are in prison – or the share of offenders counties house in state prison relative to the sum of prison and jail inmates. Namely, if, prior to AB 109, counties sent relatively more offenders to prison (rather than jail), then following Realignment, the county would become responsible for a larger percentage of realigned offenders than counties that were less reliant on state prisons. And because counties in the top third of the distribution have smaller pre-AB 109 jail population-to-capacity ratios, their county jails can absorb relatively more realigned offenders.

However, in Panel B, there is no evident relationship between the percent change in the prison incarceration rate and the percent change in jail incarceration rate. On average, the largest percent decline in the prison incarceration rate occurs in the middle third of counties. This suggests that county jails vary in their capacity to absorb realigned offenders, or that counties may have preferences for alternative sanctions like probation.

In Panel A, relative to counties in the middle and top thirds of the distribution, counties in the bottom third of the distribution are subject to more capacity constraints (the average pre-AB 109 jail population-to-capacity ratio is 93.57 percent). These counties also sentence a larger share of offenders to probation (rather than prison or jail). This does not necessarily indicate a preference for alternative sanctions. Because the jails in counties in the bottom third are relatively more capacity-constrained, they may be forced into using more probation so as not to exceed rated capacity. Furthermore, Figure 3.1 shows that the number of individuals on probation was declining prior to Realignment (possibly due to declining arrest rates, as discussed in Section 3.1). Yet, in counties in the bottom third of the percent change in the jail incarceration rate distribution, the probation population fell the least (Panel B). Therefore, because of the high, pre-AB 109 capacity utilization rates, it is possible that counties in the bottom third of the distribution substituted probation for county jail sentences. On the other hand, counties in the top third of the distribution continued to experience relatively larger declines in the probation population following AB 109. Because these counties had the capacity to absorb offenders that would have otherwise received a prison sentence, they had no need to use alternative sanctions.

Similarly, Panel B shows that counties in the bottom third of the distribution experienced the largest percent increase in releases. As in Lofstrom and Raphael (2013a), to avoid exceeding rated capacity, some counties may have had to release relatively more offenders through capacity-constrained releases (early release) or pre-trial releases.[28] Therefore, because there may be a jail crowd-out effect, the percent change in jail incarceration rates is small. On the other hand, counties in the top third of the distribution experienced a much smaller percent increase, on average, in county jail releases because they had the capacity to absorb realigned offenders.

3.6.2 Fiscal Impact

Although California provided funding to assist counties in managing an increased offender caseload, differences in the demographic, economic, and criminal justice characteristics of counties could have affected the magnitude of AB 109's fiscal impact. Like Table 3.2, Table 3.3 divides counties into three equally-sized groups according to the percent change in the jail incarceration rate following AB 109. It focuses on counties' fiscal characteristics and on the possible fiscal effects of Realignment.

Recall that the fiscal impact per resident is calculated by multiplying the average cost per jail inmate (or jail expenditures per jail inmate) by the change in the number of jail inmates following AB 109. In Table 3.3, counties in the bottom third of the distribution have the highest average cost per jail inmate. However, among those counties, the jail incarceration rate declined. The middle third of the distribution has the lowest average cost per jail inmate. And because their percent increase in the jail incarceration rate is almost 4 times smaller than that of counties in the top third of the distribution, counties in the top third of the distribution experience a much larger fiscal impact per resident - a possible \$29.77 increase per resident in jail expenditures, relative to a \$7.30 increase per resident.

Furthermore, among counties in the bottom and middle thirds of the distribution, on

[28]This is especially true if counties have a court-ordered population cap.

average, the AB 109 funds per resident (provided by the state) cover the estimated impact of AB 109. Yet, among counties in the top third of the distribution, the AB 109 funds per resident cover less than half of the estimated fiscal impact. While my estimate of fiscal impact is a rough measure of how AB 109 may have affected the county budget, this suggests that the first-year allocation of AB 109 funds may not have been distributed to the counties that needed the funds the most.[29] This is further emphasized by the fact that comparing jail expenditures per resident between Fiscal Year 2011 and Fiscal Year 2013, expenditures per resident only increased by 0.32 percent in counties in the bottom third of the distribution. Jail expenditures per resident increased by 5.84 percent, on average, among counties in the top third of the distribution.

Lastly, moving from the bottom third of the distribution to the top third, (the average of) median income declines. So, counties that experienced the largest percent increase in jail incarceration rates were the least well-off in terms of median income. To assess the relationship between median income and fiscal impact per resident, Figure 3.5 presents two maps of California's counties shaded by median income (Figure 3.5a) and shaded by the AB 109 fiscal impact per resident (Figure 3.5b). As median income increases, the shading becomes lighter. Because median income and the fiscal impact per resident are negatively correlated, as the fiscal impact per resident increases, the shading becomes darker. Therefore, if counties with the lowest level of median income are impacted the most by AB 109, then they should be shaded the same color across Figures 3.5a and 3.5b. Figure 3.5a shows considerable geographic concentration in median income. Counties in northern California have the lowest levels of median income and counties along the coast of southern and central California have the largest levels median income. While there is less geographic concentration in fiscal impact per resident, most counties below the median of median income appear to experience above median fiscal impacts.

[29]Figure 3.A.1, in Appendix 3.A, presents two maps of California's counties to illustrate whether relatively more AB 109 funds per resident were allocated to counties that experienced large (estimated) fiscal impacts. Figure 3.A.1a shades counties by the magnitude of the AB 109 fiscal impact per resident. Figure 3.A.1b shades counties by the magnitude of the AB 109 funds per resident received. As the fiscal impact per resident increases and as AB 109 funds received (per resident) increase, the shading becomes darker.

Overall, it appears that the passage of AB 109 placed a larger fiscal burden on lower-income counties and AB 109 funds may have been disproportionately allocated to relatively richer counties, who are more capacity-constrained and, therefore, may use cheaper alternatives to incarceration. Note, however, that I am only looking at the impact on jail expenditures. AB 109 also shifted state parole supervision to county probation departments, so probation caseloads and costs would have also increased. A more complete picture of the fiscal impact of AB 109 would examine how both jail and probation costs were affected.

3.6.3 Regression on the percent change in jail incarceration rates

More formally, I estimate a regression of the percent change in the jail incarceration rate on the percent change in the prison incarceration rate, conditional on pre-AB 109 county characteristics.

$$\%\Delta Jail_c = \alpha_1 + \beta_1 * \%\Delta Prison_c + \beta_2 * Capacity_c + X_c * \beta_3 + \varepsilon_c \tag{3.1}$$

Let $\%\Delta Jail_c$ represent the percent change in the jail incarceration rate following AB 109 in county c. $\%\Delta Prison_c$ represents the percent change in the prison incarceration rate following AB 109 in county c. $Capacity_c$ represents a county's pre-AB 109 jail population-to-capacity ratio. Let X_c be a vector of county-level, pre-AB 109 covariates including median income, the share of the county's population that is black or Hispanic, the share of incapacitated offenders in state prison, and the share of all offenders that are under probation. The error term is denoted by ε_c. I estimate robust standard errors.

Table 3.4 displays the coefficients from the regression on the percent change in the jail incarceration rate. Column 1 includes all variables listed in Equation 3.1 except for the jail population-to-capacity ratio. Column 2 includes the capacity ratio. In Column 1, the coefficient on the percent change in the prison incarceration rate is negative but not statistically significant.

The sign of β_1 suggests that, conditional on the pre-AB 109 characteristics in X_c, as the percent change in the prison incarceration rate falls, there is an associated increase in the average percent change in the jail incarceration rate (see Figure 3.2). However, in Column 2, the coefficient on $\%\Delta Prison_c$ becomes positive when the pre-AB 109 jail population-to-capacity ratio is included as a regressor; it remains statistically insignificant.

In Columns 1 and 2, the coefficient on median income is negative and statistically significant. This suggests that, as illustrated in Table 3.3, a decline in median income is associated with an increase in the percent change in the jail incarceration rate. Additionally, the coefficient on the percent of incapacitated offenders in prison (relative to offenders in jail and prison) is positive and statistically significant. Like Table 3.2, this illustrates that, all else equal, counties that used prison relatively more than jail, prior to the passage of AB 109, also experienced an increase in the average percent change in the jail incarceration rate.

In Column 2, the coefficient on the jail population-to-capacity ratio is negative and statistically significant - all else equal, an increase in the capacity ratio is associated with a decline in the average percent change in the jail incarceration rate. As with Table 3.2, this suggests that counties that were capacity-constrained prior to AB 109 experienced smaller (percent) changes in the jail incarceration rate. Table 3.2 also suggests that these counties likely had to increase the number of offenders released from jail or increase the use of alternative sanctions. Note that including the jail population-to-capacity ratio reverses the sign on the coefficients on the share of the county population that is black or Hispanic and the percent of all offenders that are on probation.

In all, I find that, consistent with previous literature, capacity-constrained counties experienced smaller percent increases in the jail incarceration rate (on average). In order to absorb the increased offender caseload, Table 3.2 suggests that these counties increased their use of early release and possibly probation. In contrast to previous literature, I do not find a statistically significant relationship between the percent change in the prison and jail incarceration rates following AB 109. Most notably, I find that the fiscal impact (per resident) may have been the largest for lower-income counties.

3.7 Proposition 47

Next, I study the extent to which the implementation of Proposition 47 mitigated the effects of AB 109. Figure 3.1 illustrates how California's prison, jail, and probation populations changed following the adoption of Prop 47. While there were small changes to the prison population (black dashed line), there was a discrete decline in the county jail population (black solid line).[30] Figure 3.3, which presents a scatter plot of the percent change in the jail incarceration rate following AB 109 and the percent change in jail incarceration rate following Prop 47, illustrates a weak negative correlation, such that counties with larger percent increases in the jail incarceration rate following AB 109 had larger percent declines in the jail incarceration rate following Prop 47. However, since the association is weak, Figure 3.3 suggests that other county-level characteristics might explain the percent change in the jail incarceration rate following Prop 47 and that it may not have brought cost savings to the counties most impacted by AB 109.

3.7.1 Proposition 47 and the Criminal Justice System

Section 3.6 showed that the counties which experienced the smallest percent increases in their jail incarceration rates following AB 109 were more likely to be capacity-constrained. As a result, these counties increased their use of pre-trial and early releases so that they could absorb the realigned offender population. Following the implementation of Prop 47, which was expected to increase jail space through shorter sentence lengths, all counties experienced a decline in their Prop 47 jail populations. However, due to the Prop 47-induced availability of jail space, counties with high pre-AB 109 population-to-capacity ratios may not have needed to use

[30] There is also a sharp decline in the probation population (gray dashed line). Bird et al. (2018) finds that Prop 47 sharply reduced arrests, which is strongly correlated with probation populations.

capacity-constrained releases as often. Consequently, the overall jail incarceration rate may have only changed slightly for these counties (Nguyen et al. 2016).

Table 3.5 stratifies counties into three groups based on the distribution of the percent change in the jail incarceration rate following Proposition 47. Counties in the lower third of the distribution experience a (percent) decline in the average jail incarceration rate that is twice as large as that of counties in the middle third of the distribution. Counties in the top third of the distribution experience, on average, an increase in the jail incarceration rate following the implementation of Prop 47.

While counties in the lower third of the distribution experienced larger percent increases in the jail incarceration rate following AB 109, there is little difference between the percent change in jail incarceration rates after AB 109 in the middle and top thirds of the distribution.[31] Furthermore, the average pre-AB 109 jail population-to-capacity ratio does not differ much across the three groups. Interestingly, counties that experienced the largest percent decline in the jail incarceration rate after Prop 47 had the largest average post-AB 109 (pre-Prop 47) jail population-to-capacity ratio. It is possible that the counties in the bottom third of the distribution experienced a large change in the population-to-capacity ratio because they increased their use of early releases at a smaller rate relative to counties in the top third of the distribution. Following Prop 47, counties in the bottom third of the distribution experienced larger percent declines in the use of pre-trial and early release.

Table 3.5 also shows that among counties in the bottom third of the distribution of the percent change in jail incarceration rates following Prop 47, the pre-AB 109 share of property arrests (relative to all offense types) was, on average, larger relative to that of counties in the middle and top thirds. Counties that arrest relatively more property offenders may experience larger percent decreases in the jail incarceration rate since most of the offenses re-classified by

[31] Figure 3.A.2, in Appendix 3.A, illustrates the percent change in each county's jail incarceration rate following AB 109 (Figure 3.A.2a) and following Prop 47 (Figure 3.A.2b). Prop 47 does not appear to decrease the jail incarceration rate more in counties whose jail incarceration rate increased the most (in percent) due to AB 109.

Prop 47 were low-level property offenses.

Overall, Table 3.5 does not appear to provide a clear relationship between the change in jail incarceration rates following AB 109 and Prop 47. Namely, Prop 47 does not appear to undo the effects of AB 109 in the counties most affected by Realignment. In fact, stratifying counties across the percent change in jail incarceration rates following AB 109 (as in Table 3.2), the largest percent decline in jail incarceration rates after Prop 47 (on average) was among counties that experienced decreases or small increases in jail incarceration rates after AB 109.

3.7.2 Fiscal Impact

Given that Prop 47 does not appear to undo much of the increases in jail incarceration rates due to AB 109, I would not expect Prop 47 to relieve the fiscal burden placed on counties most impacted by AB 109. Like Table 3.5, Table 3.6 stratifies counties into three groups based on the distribution of the percent change in the jail incarceration rate following Proposition 47. Note that, on average, the top third of the distribution experienced an increase in jail incarceration rates; around 25 percent of California's counties experienced an increase in the jail incarceration following Prop 47. Relative to counties in the bottom and middle thirds, these counties had relatively lower median incomes, on average.

Recall that fiscal impact is defined as the product of the pre-AB 109 average jail inmate cost (or jail expenditures per inmate) and the change in the jail population. A negative fiscal impact per resident indicates cost savings (since the number of offenders fell after Prop 47). Therefore, the bottom third of the distribution experienced the largest potential fiscal benefit from the reduced jail incarceration rate. These counties also experienced, on average, the largest fiscal impact from AB 109 and had higher median incomes. However, Table 3.3 showed that the counties that experienced the largest AB 109 fiscal impact had lower median incomes, on average.

Similarly, Figure 3.6 presents a scatter plot of the fiscal impact (per resident) of AB 109 and the fiscal impact (per resident) of Prop 47. The black circle-markers represent counties that fall below the median of median income, and the gray circle-markers represent counties that have above-median median income. The fiscal impact of Prop 47 is multiplied by -1 so that a positive value represents savings. Figure 3.6 illustrates a (weak) positive relationship between the fiscal cost of AB 109 and the fiscal savings from Prop 47. Yet, most below-median median income counties lie below the 45 degree line, indicating that among low-income counties the fiscal savings from Prop 47 did not mitigate the fiscal impact of AB 109. On the other hand, counties with above-median median income are more concentrated along the 45 degree line, suggesting that among high-income counties the savings generated by Prop 47 did a better job of easing the fiscal impact of AB 109. As a result of Table 3.6 and Figure 3.6, in the following section I include an interaction term between median income and the percent change in the jail population following AB 109 to assess whether the effect of the latter depends on median income.

3.7.3 Regression on the percent change in jail incarceration rates

More formally, I estimate a regression of the percent change in the jail incarceration rate following Prop 47 on the percent change in the jail incarceration rate following AB 109, conditional on pre-AB 109 county characteristics.

$$\begin{aligned}\%\Delta Jail_c^{Prop47} = \alpha_2 + \gamma_1 * \%\Delta Jail_c^{AB109} + \gamma_2 * Income_c + \gamma_3 * \%\Delta Jail_c^{AB109} * Income_c \\ + \gamma_4 * Capacity_c^{Prop47} + X_c^{Prop47} * \gamma_5 + u_c\end{aligned} \tag{3.2}$$

Let $\%\Delta Jail_c^{Prop47}$ represent the percent change in the jail incarceration rate following Prop 47 in county c. $\%\Delta Jail_c^{AB109}$ represents the percent change in the jail incarceration rate following AB 109 in county c. $Capacity_c^{Prop47}$ represents a county's post-AB 109, pre-Prop 47 jail population-to-capacity rate. $Income_c$ is county c's pre-AB 109 median income. Let X_c^{Prop47} be a vector of county-level, pre-AB 109 covariates including the share of the county's population that is black or Hispanic and the share of total arrests that are for property offenses. The error term is denoted by u_c. I estimate robust standard errors.

Table 3.7 presents coefficients from the regression on the percent change in the jail population following Proposition 47. In Column 1, I include the share of total arrests that are for property offenses. Prop 47 reduced several low-level property offenses from felonies to misdemeanors. Therefore, in counties in which a larger share of arrests are for property offenses, there might be larger declines in the jail incarceration rate, all else equal. Because median income is strongly (and positively) correlated with the share of property offense arrests, I omit median income from the regression in Column 1. Columns 2 and 3 include median income as a covariate (and omit the share of property arrests). Column 3 includes the interaction term between median income and the percent change in the jail population following AB 109.

Across all columns, the coefficient on the post-AB 109 jail population-to-capacity ratio is negative, and it is statistically significant in Columns 1 and 3. All else equal, as jails become more capacity-constrained the percent change in the jail incarceration rate following Prop 47 declines. This is also evident in Table 3.5, where counties in the bottom third of the distribution of the percent change in the jail incarceration rate after Prop 47 have a larger post-AB 109 population-to-capacity ratio (on average).

In Column 1, the coefficient on the share of arrests for property offenses is negative and statistically significant. As the share of property arrests increases, the percent change in the jail incarceration rate declines, suggesting that counties which arrest relatively more individuals for property offenses are more likely to be impacted by the re-classification of low-level property offenses.

In Columns 1 and 2, the coefficient on the percent change in the jail incarceration rate following AB 109 is negative; it is only statistically significant in Column 2. Therefore, as illustrated in Figure 3.3, as the percent change in the AB 109 jail incarceration rate increases, the average percent change in the Prop 47 jail incarceration rate declines. Unlike Table 3.5, the regression coefficients in Columns 1 and 2 indicate that, all else equal, Prop 47 appears to mitigate the effects of AB 109.

In Column 2, the coefficient on median income is negative and statistically significant, such that, all else equal, as median income increases the average percent change in the jail incarceration rate declines. Like Table 3.5, this suggests that Prop 47 may have impacted jail incarceration rates in counties that are better off, unlike AB 109 which imposed a relatively larger fiscal burden on counties with low median incomes.

Lastly, like Column 2, the coefficients on $\%\Delta Jail_c^{AB109}$ and $Income_c$ in Column 3 are negative and statistically significant. Furthermore, the coefficient on the interaction term, $\%\Delta Jail_c^{AB109} * Income_c$, is positive and statistically significant. Therefore, all else equal, the effect of a one percent increase in the jail incarceration rate (following AB 109) is smaller for counties with higher median incomes. Or for each thousand dollar increase in median income, counties with larger percent increases in the jail incarceration rate have smaller changes in the outcome variable.

In all, I find that, as with AB 109, the jail population-to-capacity ratio continues to be associated with how counties respond to the adoption of Prop 47. I also find that while the relationship between the percent change in jail incarceration rates following AB 109 and Prop 47 is more ambiguous in Table 3.5, the coefficients in Table 3.7 suggest that the percent change in the jail incarceration rate following AB 109 is negatively associated with the percent change in the jail incarceration rate following Prop 47, but the magnitude of the association differs across median income.

3.8 Conclusion

In 2011, the U.S. Supreme Court mandated that California reduce its prison population from 190 to 137.5 percent of design capacity. In response, California implemented a series of reforms. The first was SB 678 (July 2009), which incentivized counties to send fewer offenders to prison for probation failures. Because SB 678 did not sufficiently reduce the prison population, in October 2011, California implemented AB 109. This reform shifted the supervision responsibility for low-level offenders from the state prison system to county jails. While AB 109 successfully reduced the prison population, the number of state prison inmates did not fall below the court-mandated target. Therefore, in November 2014, California implemented Proposition 47, which re-classified low-level felonies as misdemeanors.

As a result of AB 109, counties faced an increased offender caseload, and variation in counties' demographics, economies, and use of state prison (or incarceration in general) may have contributed to variation in the extent to which county jails absorbed realigned offenders. While I do not find a statistically significant relationship between the percent change in the prison and jail incarceration rates following AB 109, I find that in order to absorb the increased offender caseload, counties whose jails were capacity-constrained, increased their use of early release and possibly probation. Most notably, I find that the fiscal impact (per resident) may have been the largest for lower-income counties and that the AB 109 funds allocated to counties fell short in covering counties? increased costs, particularly among counties that experienced the largest fiscal impact.

Next, I study the extent to which the implementation of Prop 47 mitigated the effects of AB 109. I find that the jail population-to-capacity ratio continues to be associated with how counties responded to the adoption of Prop 47. I also find that the percent change in the jail incarceration rate following AB 109 is negatively associated with the percent change in the jail incarceration rate following Prop 47, but the magnitude of the association differs across median income. While Prop 47 may have mitigated some of the effects of AB 109, I find little evidence suggesting Proposition 47 mitigated the fiscal impacts of AB 109 among low-income counties.

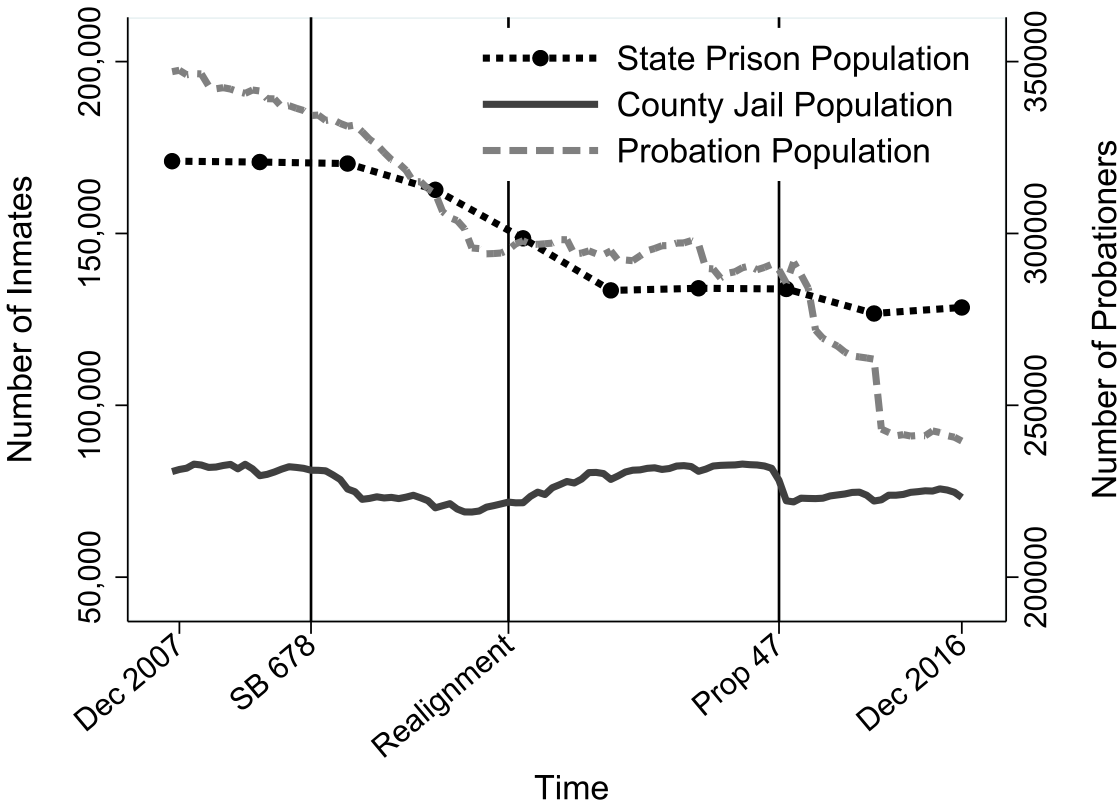


Figure 3.1: Number of offenders receiving prison, jail, or probation over time

This figure illustrates a time-series of California prison, jail, and probation populations. Each vertical line represents the date in which a given reform was enacted. SB 678 was implemented in July 2009, AB 109 (or Realignment) in October 2011, and Proposition 47 in November 2014. The black dashed line represents California's state prison population in December of each year (left y-axis), and the black solid line represents the average daily jail population in each month (left y-axis). The gray dashed line represents the number of individuals on probation in each month (right y-axis).

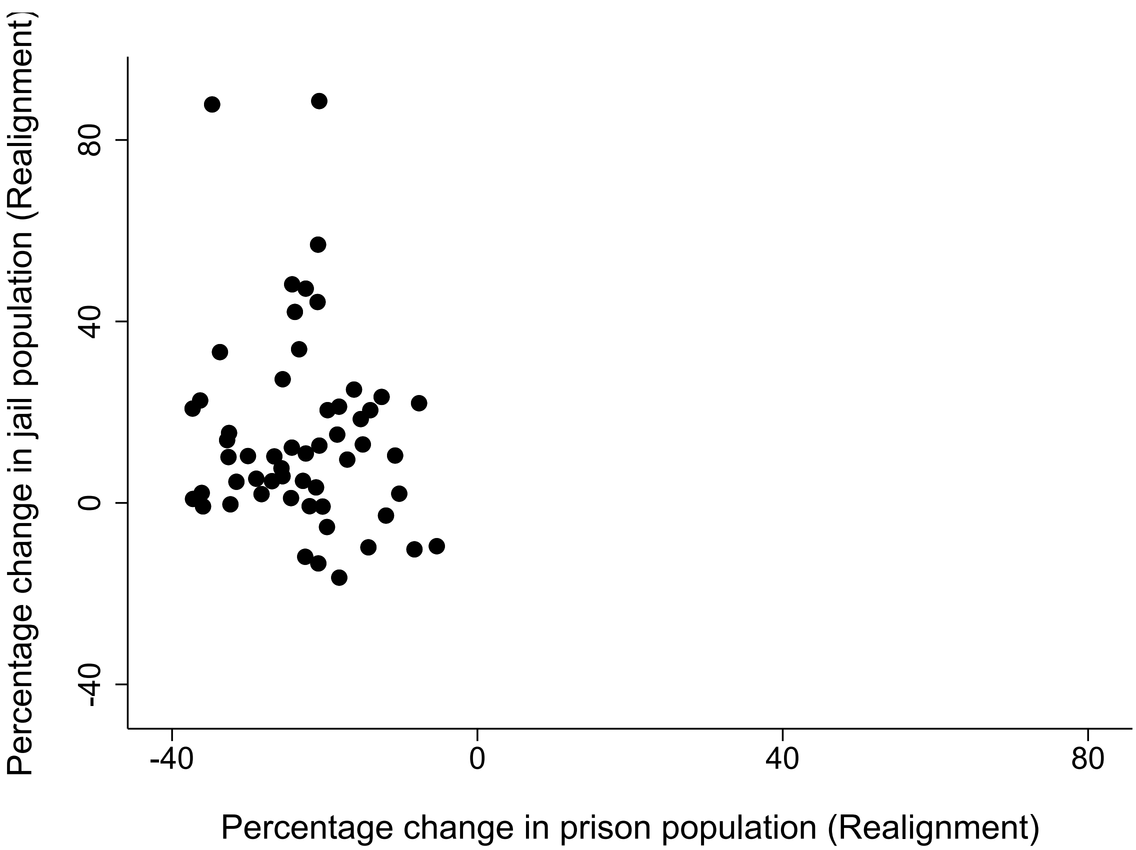


Figure 3.2: Scatter plot of the percent change in prison and jail incarceration rates following AB 109

This figure presents a scatter plot comparing the percent change in prison (x-axis) and jail (y-axis) incarceration rates following the implementation of AB 109. Each point represents a different county. The incarceration rate is defined as the number of inmates per thousand residents. Jail and prison populations as of December 2010 and December 2012 are used to calculate the percent change in jail and prison incarceration rates. The black line has a slope of -1. Counties that fall on the black line experienced equally-sized percent changes in jail and prison incarceration rates (in absolute value).

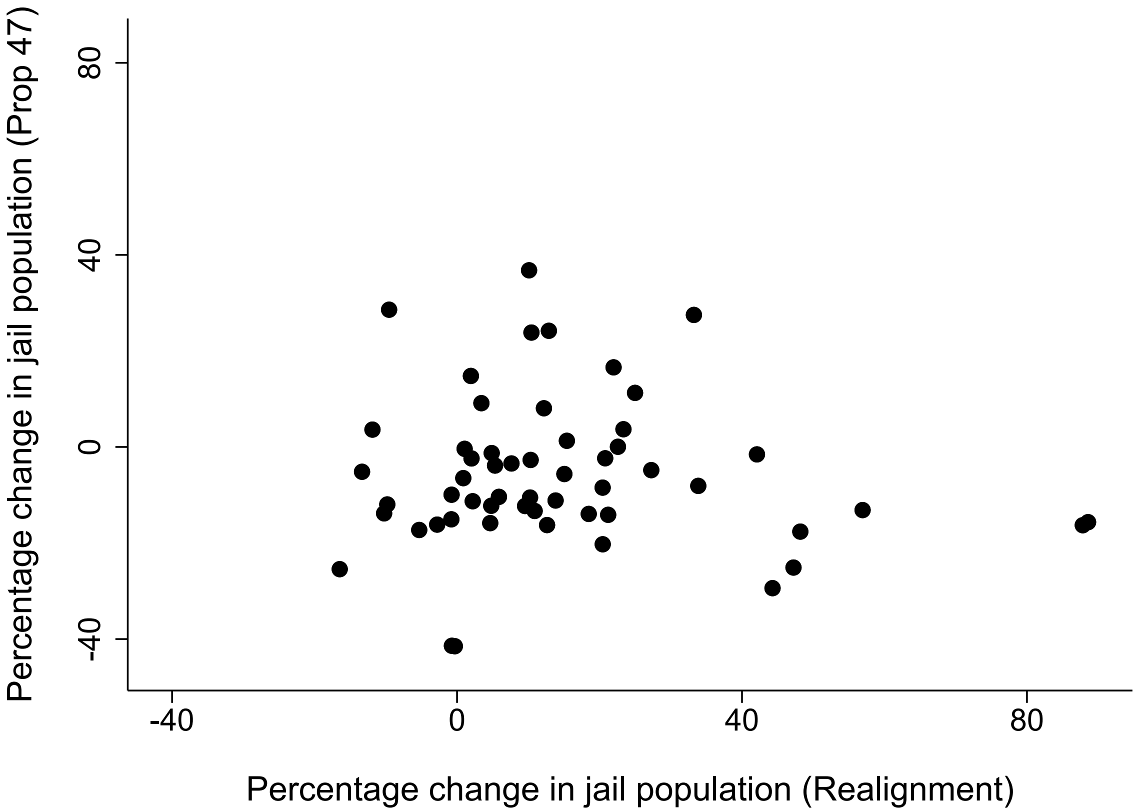


Figure 3.3: Scatter plot of the percent change in jail incarceration rates following AB 109 and Proposition 47

This figure presents a scatter plot comparing the percent change in jail incarceration rates following AB 109 (x-axis) and the percent change in jail incarceration rates following Proposition 47 (y-axis). Each point represents a different county. The incarceration rate is defined as the number of inmates per thousand residents. The percent change in the jail incarceration rate following AB 109 is calculated by comparing the average daily jail populations per thousand residents as of December 2010 and December 2012. The percent change in the jail incarceration rate following Proposition 47 is calculated by comparing the average daily jail populations per thousand residents as of December 2013 and 2015. The black line has a slope of -1. Counties that fall on the black line experienced equally-sized percent changes in the AB 109 and Prop 47 jail incarceration rates (in absolute value).

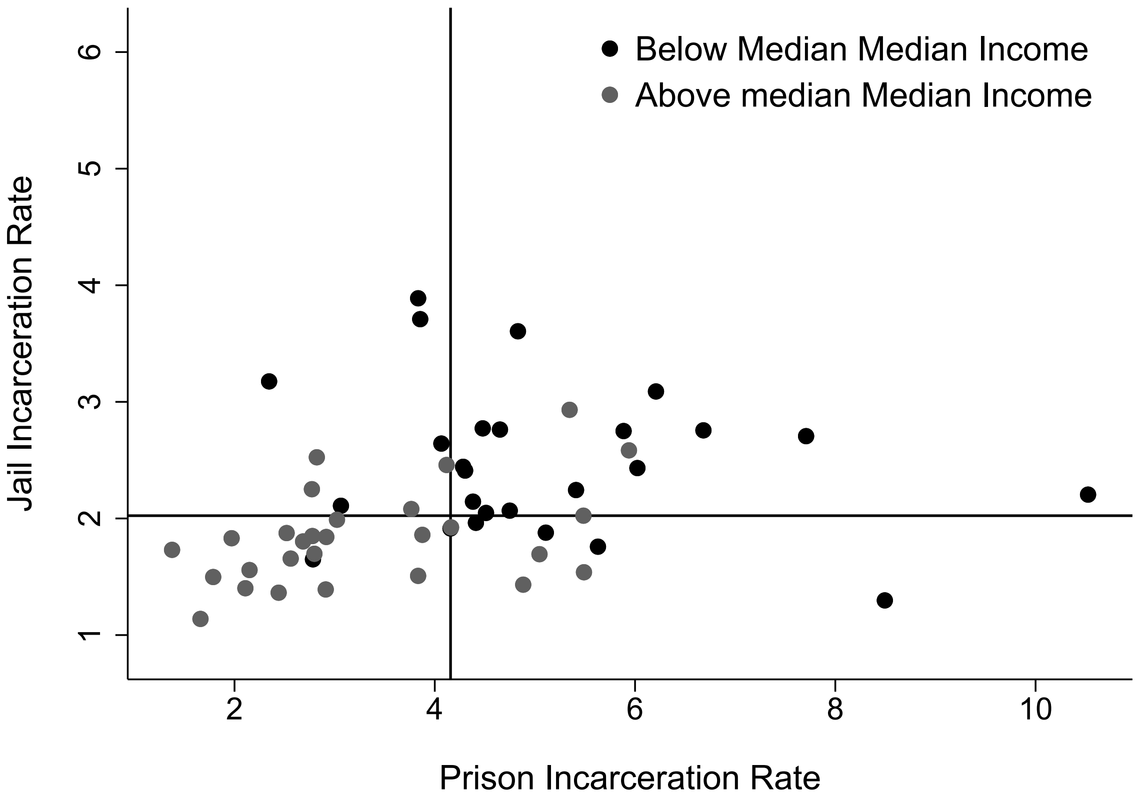
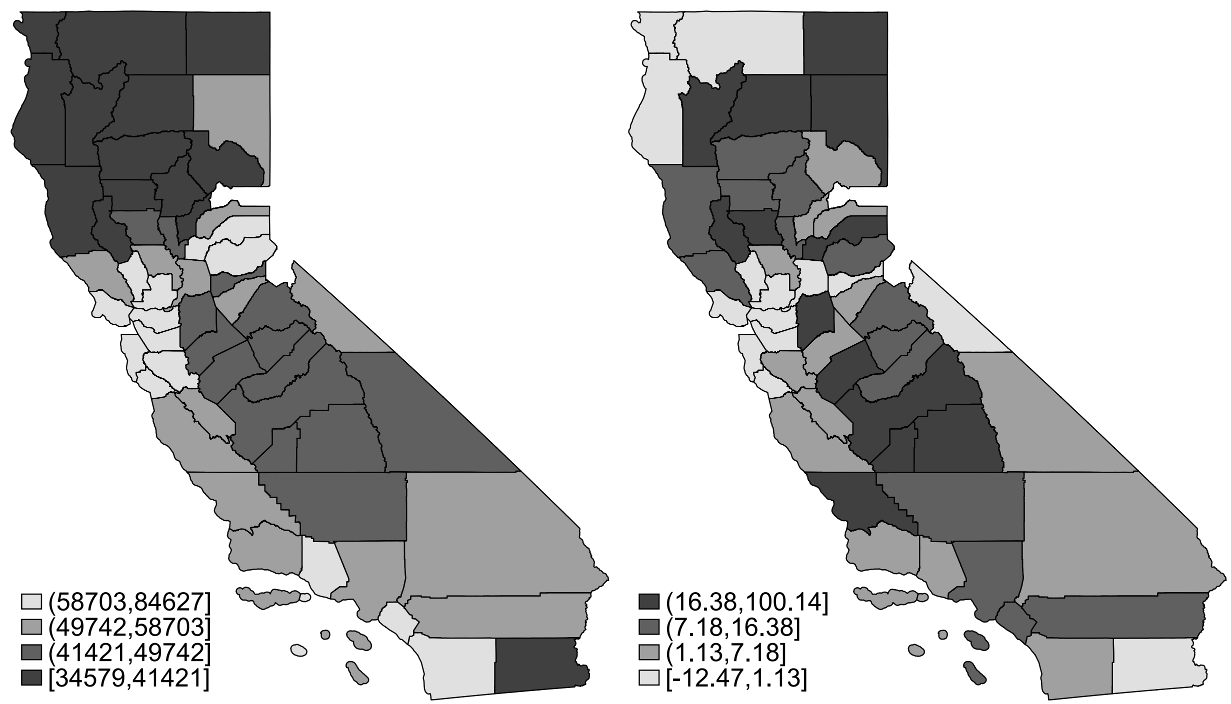


Figure 3.4: Scatter plot of counties' prison and jail incarceration rates, separated by median income

This figure presents a scatter plot of counties' pre-AB 109 prison and jail incarceration rates, defined as the number of prison and jail inmates per thousand residents. The vertical and horizontal lines represent the median prison incarceration rate and the median jail incarceration rate, respectively. The black circle-markers represent counties that fall below the median of median income, and the gray circle-markers represent counties that lie above the median of median income.



(a) Median Income

(b) Fiscal impact per resident - AB 109

Figure 3.5: Map of California counties shaded by median income and the AB 109 fiscal impact per resident

This figure presents two maps of California shaded by median income (left map) and shaded by the AB 109 fiscal impact per resident (right). Fiscal impact is defined as the pre-AB 109 average cost per jail inmate (or jail expenditures per resident) multiplied by the change in the number of jail inmates between December 2010 and December 2012. As median income increases, the shading becomes lighter. As the fiscal impact per resident increases, the shading becomes darker. (Median income and fiscal impact per resident are negatively correlated.)

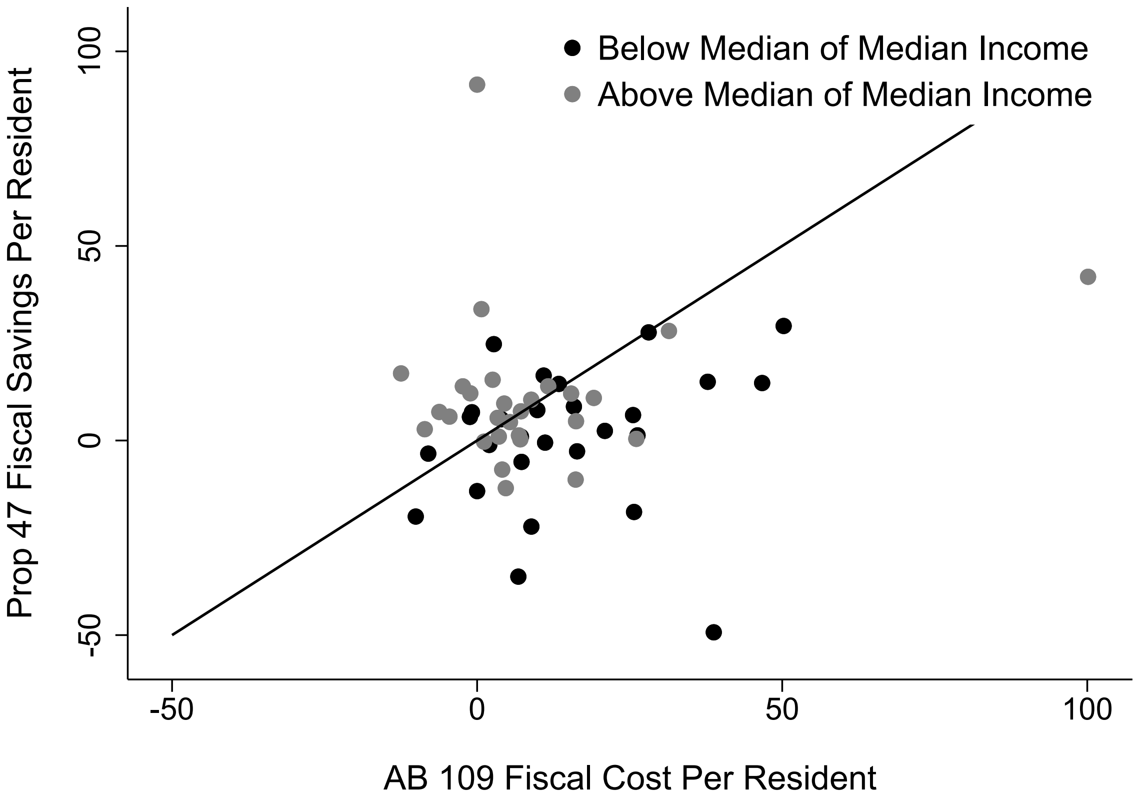


Figure 3.6: Scatter plot of the AB 109 and Proposition 47 fiscal impacts, separated by median income

This figure presents a scatter plot of the fiscal impact of AB 109 per resident (x-axis) and the fiscal impact of Prop 47 per resident (y-axis). Each point represents a different county. The AB 109 fiscal cost (or fiscal impact) is defined as the pre-AB 109 average cost per jail inmate (or jail expenditures per resident) multiplied by the change in the number of jail inmates between December 2010 and December 2012. The Prop 47 fiscal impact is defined as the pre-AB 109 average cost per jail inmate multiplied by the change in the number of jail inmates between December 2013 and December 2015. This value is then multiplied by -1 to obtain the Prop 47 fiscal savings. The black circle-markers represent counties that fall below the median of median income, and the gray circle-markers represent counties that lie above the median of median income. The black line is a 45 degree line (slope of 1).

Table 3.1: Descriptive statistics for counties by tercile of median income

	Bottom 3^{rd}	Middle 3^{rd}	Top 3^{rd}
Panel A: County Demographics			
Median Income	39,370	48,863	65,646
	(2,855)	(3,613)	(9,772)
Share of County Pop. Urban	55.64	73.54	90.00
	(27.12)	(24.81)	(11.79)
Share of County Pop. Black	1.94	4.80	4.13
	(1.38)	(3.56)	(4.23)
Share of County Pop. Hispanic	26.32	33.79	28.61
	(21.93)	(16.58)	(12.19)
Heterogeneity Index	0.54	0.67	0.72
	(0.12)	(0.22)	(0.25)
Panel B: Fiscal Measures			
County Expenditures Per Resident	1,845	1,638	1,326
	(494)	(700)	(210)
Perc. County Expend. Spent on Jails	4.91	5.56	6.16
	(1.13)	(1.27)	(2.20)
Jail Expenditures Per Jail Inmate	32,727	40,907	46,333
	(9,023)	(16,381)	(17,359)
Jail Expenditures Per Resident	88	91	81
	(22)	(44)	(28)
Panel C: Criminal Justice Indicators			
Arrests Per Thousand Residents	14.78	11.98	9.09
	(3.10)	(2.31)	(1.41)
Jail Inmates Per Thousand Residents	2.71	2.18	1.76
	(0.96)	(0.59)	(0.34)
Prison Inmates Thousand Residents	5.03	4.79	2.94
	(1.70)	(1.82)	(1.05)
Pre-AB 109 Jail Pop-to-Capacity Ratio	80.31	82.20	88.00
	(23.65)	(22.51)	(16.90)

Counties are sorted into three equally-sized groups based on the distribution of median income. The table is divided into three panels: (1) county-level demographics, (2) measures of county-level expenditures, and (3) criminal justice indicators. Characteristics in Panel A are pre-AB 109 demographics (using 2010 data). Panel B uses county expenditures from Fiscal Year 2011 (July 2010 to June 2011). Panel C uses the average daily (jail) population as of December 2010 and the prison population as of December 31, 2010.

Table 3.2: AB 109 - counties by tercile of percent change in the jail incarceration rate

	Bottom 3^{rd}	Middle 3^{rd}	Top 3^{rd}
Panel A: Criminal Justice Indicators			
Perc. Change Jail - AB 109	-3.91	9.82	37.04
	(6.28)	(3.57)	(21.39)
Perc. Incapacitated Offenders in Prison	58.55	64.81	69.27
	(7.64)	(7.25)	(8.33)
Perc. All Offenders Under Probation	67.70	62.87	65.60
	(12.79)	(15.24)	(13.06)
Pre-AB 109 Jail Pop-to-Capacity Ratio	93.57	83.39	74.31
	(22.19)	(17.64)	(19.39)
Panel B: Changes in Sanctions			
Perc. Change Prison - AB 109	-21.61	-24.71	-22.50
	(9.60)	(6.44)	(8.24)
Perc. Change Probation - AB 109	-3.37	-6.49	-14.26
	(38.72)	(21.93)	(22.36)
Perc. Change Releases - AB 109	87.96	47.88	16.61
	(158.22)	(134.69)	(50.85)

Counties are sorted into three groups according to the distribution of the percent change in jail incarceration rates between December 2010 and December 2012. The table is divided into two panels. Excluding the first row, Panel A lists descriptive statistics for counties' pre-AB 109 criminal justice characteristics (using 2010 data). Panel B lists descriptive statistics for the percent change in the prison incarceration rate, the percent change in probation population per thousand residents, and the percent change in pre-trial and capacity-constrained (early) releases.

Table 3.3: AB 109's fiscal impact - counties by tercile of percent change in the jail incarceration rate

	Bottom 3^{rd}	Middle 3^{rd}	Top 3^{rd}
Perc. Change Jail - AB 109	-3.91	9.82	37.04
	(6.28)	(3.57)	(21.39)
Jail Expenditures Per Jail Inmate	42,866	36,304	41,089
	(17,118)	(15,731)	(13,932)
Fiscal Impact Per Resident - AB 109	-2.21	7.30	29.77
	(4.99)	(3.44)	(20.32)
AB 109 Fund Allocations Per Resident	8.79	10.59	11.38
	(2.91)	(3.55)	(3.04)
Perc. Change Jail Expend. Per Resident - AB 109	0.32	3.87	5.84
	(7.72)	(7.81)	(17.70)
Jail Expenditures Per Resident	99	77	83
	(38)	(31)	(25)
Median Income	56,364	51,560	46,991
	(16,370)	(10,359)	(9,015)

Counties are sorted into three groups according to the distribution of the percent change in jail incarceration rates between December 2010 and December 2012. All characteristics, except for the percent change in the jail incarceration rate and the percent change in jail expenditures per resident, are pre-AB 109. County expenditures are from Fiscal Year 2011 (July 2010 - July 2011). Fiscal impact is defined as the average cost per jail inmate (or jail expenditures per resident) multiplied by the change in jail inmates between December 2010 and December 2012.

Table 3.4: Regression on the percent change in the jail incarceration rate - AB 109

	(1)	(2)
Perc. Change Prison - AB 109	-0.062	0.202
	(0.268)	(0.252)
Median Income (in thousands)	-0.473**	-0.323*
	(0.199)	(0.182)
Share of County Pop. Black	0.018	-0.137
	(0.945)	(0.756)
Share of County Pop. Hispanic	-0.025	0.162
	(0.151)	(0.113)
Perc. Incapacitated Offenders in Prison	0.998***	0.823***
	(0.211)	(0.216)
Perc. All Offenders Under Probation	0.211	-0.018
	(0.203)	(0.182)
Pre-AB 109 Jail Pop-to-Capacity Ratio		-0.530***
		(0.159)
Mean Pct. Chg in Jail	14.73	14.73
N	55	55

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This table lists coefficients from a regression on the change in the jail incarceration rate following the implementation of AB 109 (see Equation 3.1). County-level, pre-AB 109 covariates include median income, the share of the county's population that is black or Hispanic, the share of incapacitated offenders in state prison (relative to the sum of offenders in jail and in prison), and the share of all offenders that are under probation (relative to the sum of offenders in jail and prison, and on probation). Column 2 includes the jail population-to-capacity ratio as a regressor. Robust standard errors are in parentheses.

Table 3.5: Proposition 47 - counties by tercile of percent change in the jail incarceration rate

	Bottom 3^{rd}	Middle 3^{rd}	Top 3^{rd}
Perc. Change Jail - Prop 47	-20.47	-8.42	10.61
	(8.84)	(3.47)	(12.38)
Perc. Change Jail - AB 109	20.44	11.13	12.74
	(31.14)	(16.20)	(13.81)
Pre-AB 109 Jail Pop-to-Capacity Ratio	84.79	82.27	83.69
	(22.92)	(16.85)	(23.62)
Post-AB 109 Jail Pop-to-Capacity Ratio	105.59	94.04	89.62
	(25.44)	(13.60)	(16.25)
Perc. Change Releases - AB 109	36.26	40.21	66.41
	(103.56)	(105.10)	(153.42)
Perc. Change Releases - Prop 47	-39.21	-37.30	-21.13
	(81.12)	(46.05)	(108.90)
Share of Arrests - Property	23.42	22.78	21.15
	(3.86)	(4.75)	(4.24)

Counties are sorted into three groups according to the distribution of the percent change in jail incarceration rates between December 2013 and December 2015. The percent change in the jail incarceration rate following AB 109 is estimated using the average daily jail population (per thousand residents) in December 2010 and December 2012. The percent change in the jail incarceration rate following Prop 47 is estimated using the average daily jail population (per thousand residents) in December 2013 and December 2015. The pre-AB 109 jail to population-to-capacity ratio takes the ratio of the December 2010 average daily jail population and the rated capacity as of July 2010. The post-AB 109 (pre-Prop 47) jail population-to-capacity ratio takes the ratio of the December 2013 average daily population and the July 2013 rated capacity.

Table 3.6: Proposition 47's fiscal impact - counties by tercile of percent change in the jail incarceration rate

	Bottom 3^{rd}	Middle 3^{rd}	Top 3^{rd}
Perc. Change Jail - Prop 47	-20.47	-8.42	10.61
	(8.84)	(3.47)	(12.38)
Fiscal Impact Per Resident - Prop 47	-22.11	-7.93	10.39
	(19.99)	(6.07)	(13.61)
Fiscal Impact Per Resident - AB 109	16.61	9.47	9.89
	(26.48)	(12.84)	(12.43)
Median Income	57,369	49,354	48,129
	(14,299)	(10,220)	(11,749)

Counties are sorted into three groups according to the distribution of the percent change in jail incarceration rates between December 2013 and December 2015. County expenditures are from Fiscal Year 2011 (July 2010 - July 2011). The fiscal impact due to AB 109 is defined as the average cost per jail inmate (or jail expenditures per resident) multiplied by the change in jail inmates between December 2010 and December 2012. The fiscal impact due to Proposition 47 is defined as the pre-AB 109 average cost per jail inmate (or jail expenditures per resident) multiplied by the change in jail inmates between December 2013 and December 2015.

Table 3.7: Regression on the percent change in the jail incarceration rate - Proposition 47

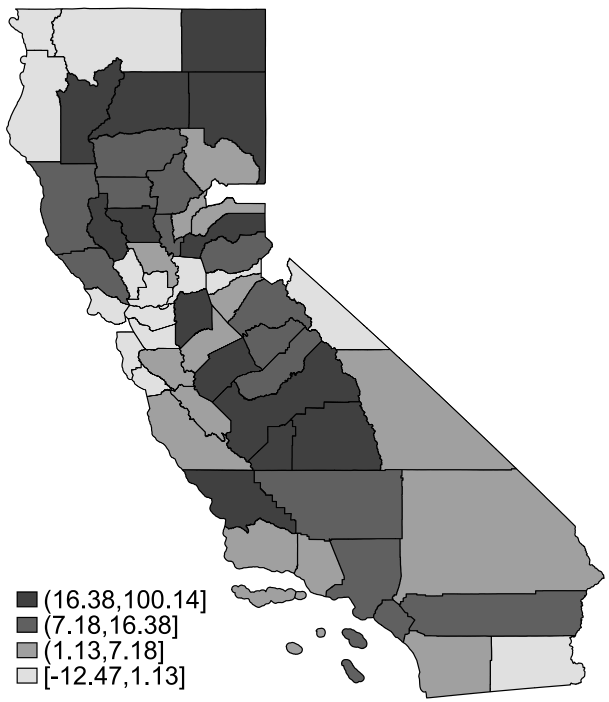
	(1)	(2)	(3)
Perc. Change Jail - AB 109	-0.137	-0.151*	-0.823***
	(0.083)	(0.085)	(0.274)
Share of Pop. Black	-35.992	-73.023	-72.967
	(73.917)	(62.241)	(53.795)
Share of Pop. Hispanic	10.890	-5.069	-5.737
	(11.758)	(13.201)	(13.008)
Share of Arrests - Property	-1.594**		
	(0.643)		
Post-AB 109 Pop-to-Capacity Ratio	-0.236**	-0.153	-0.185*
	(0.102)	(0.109)	(0.107)
Median Income (in thousands)		-0.404***	-0.446***
		(0.144)	(0.148)
Perc. Change Jail AB 109 * Median Income			0.015**
			(0.006)
Mean Pct. Chg in Jail	-5.79	-5.79	-5.79
N	55	55	55

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

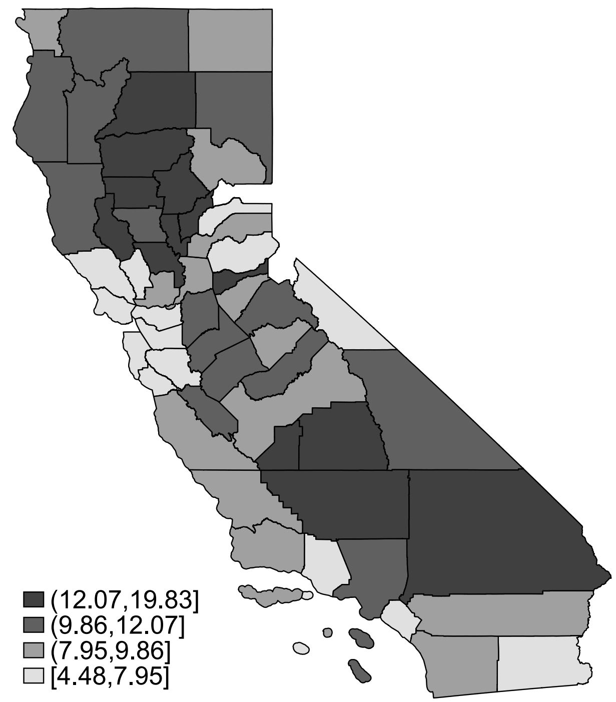
This table lists coefficients from a regression on the percent change in the jail incarceration rate following the implementation of Proposition 47 (see Equation 3.2). County-level, pre-AB 109 covariates include the share of the county's population that is black or Hispanic, and either median income or the share of total arrests that are for property offenses. Column 1 includes the share of arrests for property offenses. Columns 2 and 3 include median income. Column 3 includes a the jail population-to-capacity ratio as a regressor. Robust standard errors are in parentheses.

Appendix

3.A Additional Figures



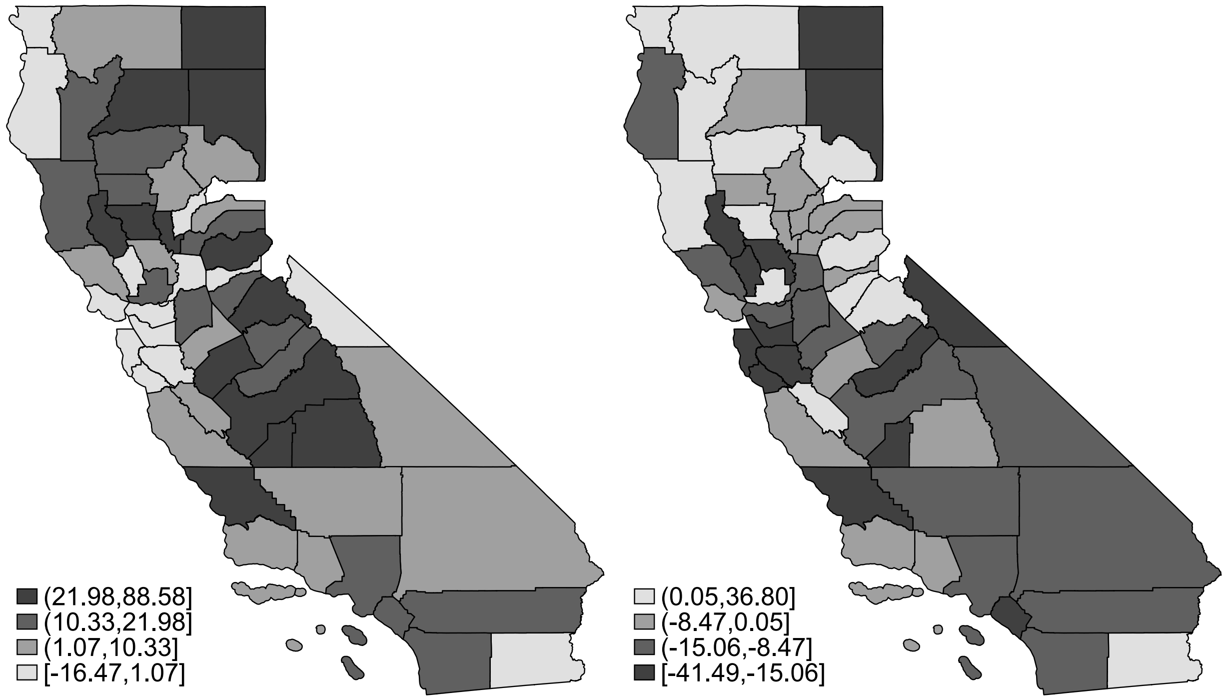
(a) Fiscal impact per resident - AB 109



(b) AB 109 funds per resident, first-year allocation

Figure 3.A.1: Map of California counties by fiscal measures

This figure presents two maps of California shaded by the AB 109 fiscal impact per resident (left) and the first-year allocation of AB 109 funds per resident (right). Fiscal impact is defined as the pre-AB 109 average cost per jail inmate (or jail expenditures per resident) multiplied by the change in jail inmates between December 2010 and December 2012. As the fiscal impact per resident increases and as AB 109 funds received (per resident) increase, the shading becomes darker.



(a) Percent change in jail incarceration rate - AB 109 (b) Percent change in jail incarceration rate - Prop 47

Figure 3.A.2: Map of California counties shaded by the percent change in jail incarceration rates

This figure presents two maps of California. The left map is shaded by the percent change in the jail incarceration rate following AB 109 (December 2010 relative to December 2012). The right map is shaded by the percent change in the jail incarceration rate following Proposition 47 (December 2013 to December 2015). The jail incarceration rate is defined as the average daily jail population per thousand residents. Because the two jail incarceration rates are negatively correlated, as the percent change in the AB 109 jail incarceration rate increases, the shading in the left map becomes darker; as the percent change in the Prop 47 jail incarceration rate decreases, the shading in the right map becomes darker.

References

Alabama Sentencing Commission. 2005. *A rational sentencing plan - ready for approval: 2005 Annual Report*.

— . 2006. *Addressing the crisis: Charting the course for reform*.

— . 2013a. *Alabama Sentencing Commission 2013 Report*.

— . 2017. *Alabama Sentencing Commission 2017 Report*.

— . 2013b. *Presumptive and Voluntary Sentencing Standards Manual*.

— . 2014. *Progress continues: Presumptive Sentencing Standards: 2014 Report*.

Alschuler, Albert W. 1968. “The prosecutor’s role in plea bargaining”. *University of Chicago Law Review* 36:50–112.

Austin, James, and Robert Lawson. 1998. “Assessment of California parole violations and recommended intermediate programs and policies”. San Francisco: National Council on Crime and Delinquency.

Autor, David H, Lawrence F Katz, and Melissa S Kearney. 2005. “Trends in US Wage Inequality: Responding to the Revisionists”. *NBER Working Paper* 11628.

Bar, Michael, Seik Kim, and Oksana Leukhina. 2015. “Gender Wage Gap Accounting: The Role of Selection Bias”. *Demography* 52 (5): 1729–1750.

Bartos, Bradley J, and Charis E Kubrin. 2018. “Can we downsize our prisons and jails without compromising public safety? Findings from California’s Prop 47”. *Criminology & Public Policy* 17 (3): 693–715.

Baum-Snow, Nathaniel, and Derek Neal. 2009. “Mismeasurement of usual hours worked in the census and ACS”. *Economics Letters* 102 (1): 39–41.

Beck, Allen J. 2006. “The importance of successful reentry to jail population growth”. Urban Institute Jail Reentry Roundtable, Washington, DC.

Bird, Mia, and Ryken Grattet. 2017. “Policy change and recidivism: The effects of Californias realignment and local implementation strategies on rearrest and reconviction”. *Criminal Justice Policy Review* 28 (6): 601–623.

Bird, Mia, and Joseph Hayes. 2013. “Funding public safety realignment”. *Public Policy Institute of California*.

Bird, Mia, Magnus Lofstrom, Brandon Martin, Steven Raphael, and Viet Nguyen. 2018. "The impact of proposition 47 on crime and recidivism". *Public Policy Institute of California*.

Blau, Francine D, and Andrea H Beller. 1988. "Trends in earnings differentials by gender, 1971–1981". *ILR Review* 41 (4): 513–529.

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations". *Journal of Economic Literature* 55 (3): 789–865.

— . 2006. "The US gender pay gap in the 1990s: Slowing convergence". *Industrial & Labor Relations Review* 60 (1): 45–66.

Britt, Chester L. 2009. "Modeling the Distribution of Sentence Length Decisions Under a Guidelines System: An Application of Quantile Regression Models". *Journal of Quantitative Criminology* 25 (4): 341–370.

California Budget & Policy Center. 2017. *Technical Appendix: Calculating County Spending on Incarceration and Responding to Crime*.

Chamberlain, Gary. 1986. "Asymptotic Efficiency in Semiparametric Models with Censoring". *Journal of Econometrics* 32:189–218.

Council of Economic Advisers. 2016. *The Long-term Decline In Prime-Age Male Labor Force Participation*. Tech. rep.

Dixon, Jo. 1995. "The Organizational Context of Criminal Sentencing". *American Journal of Sociology* 100 (5): 1157–1198.

Edwards, Griffin Sims, Stephen Rushin, and Joseph A Colquitt. 2019. "The Effects of Voluntary and Presumptive Sentencing Guidelines". *Texas Law Review, Forthcoming*.

Engen, Rodney L, and Randy R Gainey. 2000. "Modeling the effects of legally relevant and extralegal factors under sentencing guidelines: The rules have changed". *Criminology* 38 (4): 1207–1230.

Flood, Sarah, Miriam King, Steven Ruggles, and Robert J. Warren. 2015. *Integrated Public Use Microdata Series, Current Population Survey: Version 4.0. [Machine-readable database]*. Minneapolis: University of Minnesota.

Grattet, Ryken, Mia Bird, Viet Nguyen, and Sonya Tafoya. 2017. "California jails under realignment and proposition 47". *California Journal of Politics and Policy* 9 (3).

Gronau, Reuben. 1974. "Wage Comparisons –A Selectivity Bias". *Journal Of Political Economy* 82:1119–1143.

Hare, Jay. 2013. *New senetncing guidelines to impact non-violent offenders*. Ed. by Dothan Eagle. https://www.dothaneagle.com/news/new-sentencing-guidelines-to-impact-non-violent-offenders/article_4d5c3192-17ff-11e3-b9b3-0019bb30f31a.html. Accecessed: 7-10-2019.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error". *Econometrica* 47:153–161.

— . 1974. "Shadow prices, market wages, and labor supply". *Econometrica: journal of the econometric society*: 679–694.

—. 1990. “Varieties of Selection Bias”. *American Economic Review Papers and Proceedings* 80:313–318.

Herrmann, Mariesa, and Cecilia Machado. 2012. “Patterns of Selection in Labor Market Participation”. In *11th IZA/SOLE Transatlantic Meeting of Labor Economists*.

Jacobsen, Joyce, Melanie Khamis, and Mutlu Yuksel. 2015. “Convergences in Mens and Womens Life Patterns: Lifetime Work, Lifetime Earnings, and Human Capital Investment”. In *Gender convergence in the labor market*, 1–33. Emerald Group Publishing Limited.

Kirby, Brendan. 2013. *Mobile-area prosecutors rip sentencing rules calling for little or no prison time*. https://www.al.com/live/2013/10/mobile-area_prosecutors_rip_se.html. Accessed: 7-10-2019.

Lawrence, Sarah. 2014. *Court-Ordered Population Caps in California County Jails*.

Lawson, Brian. 2013. *More Alabama nonviolent offenders may avoid prison under law now in effect, DAs not happy*. Ed. by AL.com. https://www.al.com/wire/2013/10/less_prison_time_now_the_law_f.html. Accessed: 7-10-2019.

Lin, Jeffrey, and Joan Petersilia. 2014. “Follow the money: How California counties are spending their public safety realignment funds”. *Available at SSRN 2395527*.

Lofstrom, Magnus, and Steven Raphael. 2013a. “Impact of realignment on county jail populations”. *Public Policy Institute of California. Available at www. ppic. org/main/publication. asp*.

—. 2016. “Incarceration and crime: Evidence from Californias public safety realignment reform”. *The ANNALS of the American Academy of Political and Social Science* 664 (1): 196–220.

—. 2013b. *Public Safety Realignment and Crime Rates in California*.

Males, Mike, and Lizzie Buchen. 2014. *Proposition 47: Estimating local savings and jail population reductions*.

Marvell, Thomas B. 1994. “Sentencing Guidelines and Prison Population Grouth”. *J. Crim. L. & Criminology* 85:696–709.

—. 1995. “Sentencing Guidelines and Prison Population Grouth”. *Journal of Criminal Law and Criminology* 85 (3): 696–709.

McCoy, Candace. 1984. “Determinate sentencing, plea bargaining bans, and hydraulic discretion in California”. *Justice System Journal* 9:256–275.

Miethe, Terance D. 1987. “Charging and plea bargaining practices under determinate sentencing: An investigation of the hydraulic displacement of discretion”. *J. Crim. L. & Criminology* 78:155–176.

Miethe, Terance D, and Charles A Moore. 1985. “Socioeconomic disparities under determinate sentencing systems: A comparison of preguideline and postguideline practices in Minnesota”. *Criminology* 23 (2): 337–363.

Mitchell, Kelly Lyn. 2017. *Sentencing commissions and guidelines by the numbers: Cross-jurisdictional comparisons made easy by the Sentencing Guidelines Resource Center*. Tech. rep. Robina Institute in Criminal Law and Criminal Justice.

Mitchell, Ojmarrh. 2005. "A Meta-Analysis of Race and Sentencing Research: Explaining the Inconsistencies". *Journal of Quantitative Criminology* 21 (4): 439–466.

Moore, Charles A., and Terance D. Miethe. 1986. "Regulated and Unregulated Sentencing Decisions: An Analysis of First-Year Practices under Minnesota's Felony Sentencing Guidelines". *Law and Society Review* 20 (2): 253–277.

Mulligan, Casey B, and Yona Rubinstein. 2008. "Selection, investment, and women's relative wages over time". *The Quarterly Journal of Economics*: 1061–1110.

Nguyen, Viet, Mia Bird, Sonya Tafoya, and Ryken Grattet. 2016. "How has proposition 47 affected Californias jail population?" *Public Policy Institute of California*.

Nicholson-Crotty, Sean. 2004. "The Impact of Sentencing Guidelines on State-Level Sanctions: An Analysis Over Time". *Crime & Delinquency* 50 (3): 395–411.

Petersilia, Joan. 2014. "California prison downsizing and its impact on local criminal justice systems". *Harvard Law & Policy Review* 8:327–358.

Petersilia, Joan, and Jessica Snyder. 2013. "Looking past the hype: 10 questions everyone should ask about California's prison Realignment". *California Journal of Politics and Policy* 5 (2): 266–306.

Pfaff, John F. 2006. "The continued vitality of structured sentencing following Blakely: The effectiveness of voluntary guidelines". *UCLA Law Review* 54:235–307.

Piehl, Anne Morrison, and Shawn D Bushway. 2007. "Measuring and explaining charge bargaining". *Journal of Quantitative Criminology* 23 (2): 105–125.

Romano, Michael. 2015. "Proposition 47 Progress Report: Year One Implementation". Stanford Law School, Stanford Justic Advocacy Project.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. *Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database].* Minneapolis: University of Minnesota.

Schafgans, Marcia M.A., and Donald W.K. Andrews. 1998. "Semiparametric Estimation of the Intercept of a Sample Selection Model". *Review of Economic Studies* 65:497–517.

Schafgans, Marcia M.A., and Victoria Zinde-Walsh. 2002. "On Intercept Estimation in the Sample Selection Model". *Econometric Theory* 18:40–50.

Schwiebert, Jörg. 2012. "Revisiting the Composition of the Female Workforce-A Heckman Selection Model with Endogeneity". *Diskussionspapiere der Wirtschaftswissenschaftlichen Fakultt der Leibniz Universitt Hannover*.

Sorensen, Jon, and Don Stemen. 2002. "The Effect of State Sentencing Policies on Incarceration Rates". *Crime & Delinquency* 48 (3): 456–475.

Sundt, Jody, Emily J Salisbury, and Mark G Harmon. 2016. “Is downsizing prisons dangerous? The effect of California’s Realignment Act on public safety”. *Criminology & Public Policy* 15 (2): 315–341.

Taylor, Mac. 2015. *The 2015-16 Budget: Implementation of Proposition 47*. Legislative Analyst’s Office.

Turner, Susan, Terry Fain, and Shirley Hunt. 2015. *Public safety realignment in twelve California counties*. Rand Corporation.

Verma, Anjuli. 2016. “A turning point in mass incarceration? Local imprisonment trajectories and decarceration under Californias realignment”. *The Annals of the American Academy of Political and Social Science* 664 (1): 108–135.

Williams, Richard. 2016. “Using Stata’s Margins Command to Estimateand Interpret Adjusted Predictions and Marginal Effects”. Presentation.

—. 2012. “Using the margins command to estimate and interpret adjusted predictions and marginal effects”. *Stata Journal* (College Station, TX) 12 (2): 308–331.

Wooldredge, John, and Timothy Griffin. 2005. “Displaced discretion under Ohio sentencing guidelines”. *Journal of Criminal Justice* 33 (4): 301–316.