Essays on the Labor Market, Public Policy, and Economic Opportunity

by

Brenden Timpe

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2019

Doctoral Committee:

Professor Martha J. Bailey, Chair
Professor John Bound
Professor Gábor Kézdi
Assistant Professor Sarah Miller

Brenden Timpe

btimpe@umich.edu

ORCID ID 0000-0002-3547-683X

For Amy and Magdalena

ACKNOWLEDGEMENTS

Although my name is featured on the title page, this dissertation bears the fingerprints of a multitude of advisors, family members, and friends, each of whom provided invaluable guidance and encouragement.

I am extraordinarily grateful for the advice and patience of my dissertation committee: My committee chair, Martha J. Bailey; my Population Studies Center mentor, John Bound; and Gábor Kézdi and Sarah Miller. My work has benefited immensely from their incisive feedback, and I've been inspired by the examples they set as researchers and mentors.

In addition to my committee, I have learned a great deal from the numerous faculty members I encountered in courses and seminars at Michigan, including Hoyt Bleakley, Charlie Brown, John DiNardo, Sara Heller, James Hines, David Lam, Helen Levy, Elyce J. Rotella, Mike Mueller-Smith, Joel Slemrod, Jeff Smith, Mel Stephens, Adam Stevenson, Dmitriy Stolyarov, and Ugo Troiano. A large number of my fellow students at Michigan also deserve thanks for their participation in countless conversations, presentations, and reading groups, including but not limited to Luis Alejos Marroquin, Luis Baldomero Quintana, Jacob Bastian, Ariel Binder, Thomas Bridges, Giacomo Brusco, Avery Calkins, Eric Chyn, Connor Cole, Pieter de Vlieger, Xing Guo, Amelia Hawkins, Meera Mahadevan, Parag Mahajan, Dhiren Patki, Max Risch, Daniel Schaffa, Bryan Stuart, Shuqiao Sun, Evan Taylor, Tejaswi Velayudhan, and Mike Zabek. I am also grateful to J. Clint Carter, Joelle Abramowitz, Maggie Levenstein, and Lori Reeder for their help accessing restricted Census Bureau data.

Few experiences have been as educational as my collaboration with the diverse and talented group of co-authors I have enjoyed over the last several years. I could not imagine a better way to learn social science than working with Martha Bailey, Shuqiao Sun, Arline T. Geronimus, John Bound, Javier M. Rodriguez, Timothy M. Waidmann, Sarah Hamersma, Matthew Kim, Ariel J. Binder, and Avery Calkins.

I am also deeply indebted for the help I have received from the staff at the Department of Economics and the Population Studies Center at the University of Michigan, as well as the opportunity to participate in the Population Studies Center's predoctoral trainee program through an NIA training grant (T32 AG000221). I especially would like to thank Mary Braun, Laura Flak, Jennifer Garrett, Julie Heintz, Heather McFarland, and Miriam Rahl.

The seed of inspiration to pursue a Ph.D. was planted by my father many years ago and grew thanks to the encouragement he and my mother provided to explore the world around us. I am also grateful beyond words for the example set by my sisters, Miranda, Genaya, Alexandra, and Chrisanne, who, along with our parents and their spouses and children, are among my favorite people on the planet. Chad Syverson deserves special thanks for not only abiding but even encouraging the addition of another economist to the family. Finally, I was fortunate to gain via marriage a second set of impressive and supportive parents, siblings, and nieces and nephews, and I deeply appreciate their embrace of our choice to pursue life in Michigan and beyond.

Of course, I am most grateful of all for the love and encouragement of my wife, Amy, and my daughter, Magdalena. They are the strongest people I know, an inspiration for my work, and unfailingly the best part of my day.

TABLE OF CONTENTS

DEDICATION . . . ii

ACKNOWLEDGEMENTS . . . iii

LIST OF FIGURES . . . viii

LIST OF TABLES . . . ix

LIST OF APPENDICES . . . xi

ABSTRACT . . . xii

CHAPTER

I. The Long-Run Effects of America's First Paid Maternity Leave Policy . . . 1

1.1 Introduction . . . 1
1.2 The creation of America's first paid maternity leave policy . . . 4
1.2.1 The role of state disability laws and anti-discrimination policy . . . 5
1.2.2 Characteristics of STDI maternity coverage . . . 6
1.3 Expected effects of paid maternity leave . . . 8
1.3.1 Short-run effects on leave-taking and labor supply . . . 8
1.3.2 Effects on labor demand and supply . . . 9
1.3.3 Effects on children . . . 13
1.4 Data and research design . . . 15
1.4.1 Research design . . . 17
1.4.2 Take-up of STDI maternity benefits . . . 19
1.4.3 Internal validity of the research design . . . 21
1.5 Effects on women's employment and wages . . . 23
1.5.1 Heterogeneity and robustness of labor-market effects . . . 26
1.5.2 Interpretation of labor-market effects . . . 28
1.6 Effects on children . . . 29
1.6.1 Did paid leave affect fertility patterns? . . . 29

1.6.2 Effects on children in the long run 31
1.6.3 Reconciling long-run estimates with previous literature 34
1.7 Conclusion 35

II. Paid Maternity Leave and the Gender Wage Gap 53

2.1 Introduction 53
2.2 Background: Disability insurance, maternity leave, and the gender wage gap 56
2.2.1 The convergence of female and male labor-market outcomes 57
2.2.2 The role of family-friendly policies 61
2.2.3 What did maternity leave cost? 64
2.3 Data and Empirical Strategy 66
2.4 Results 68
2.4.1 The effect of STDI benefits on the gender wage ratio 69
2.4.2 The effects of STDI maternity benefits on labor supply and women's occupational distribution 70
2.5 Conclusion 73

III. Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency . . . 82

3.1 Introduction 82
3.2 The Launch of Head Start in the 1960s and Expected Effects 85
3.2.1 A Brief History of Head Start's Launch 86
3.2.2 Head Start's Mission 87
3.3 Literature Regarding the Long-Term Effects of Head Start . . 89
3.4 Data and Research Design 90
3.4.1 Measuring Exposure to Head Start 92
3.4.2 Event-Study Regression Model 92
3.4.3 Expected Effects of Exposure to Head Start by Age at Launch 93
3.4.4 Spline Summary Specification 95
3.5 Tests of Identifying Assumptions 96
3.5.1 How Much Did Head Start Increase Preschool Enrollment? 96
3.5.2 Did Head Start's Launch Correspond to Other Policy Changes? 98
3.6 Head Start's Effects on Human Capital 99
3.7 Head Start's Effects on Economic Self-Sufficiency 103
3.8 Heterogeneity in Head Start's Long-Run Effect 105
3.9 New Evidence on the Long-Term Returns to Head Start . . . 108

APPENDICES 123

A.1 STDI coverage and the implementation of anti-pregnancy discrimination laws 124
A.1.1 An additional test of the research design 127
A.2 Relevance of STDI benefits 128
A.2.1 Evidence from the decennnial Census 128
A.2.2 Evidence from the 1984-1989 SIPP 130
A.3 Maternity leave and the results of Gruber (1994) 132
A.4 Figures and Tables 134
B.1 Census Data 142
B.2 SSA's Numident: Data on County and Date of Birth 143
B.3 Data on School Age Entry Cutoffs 143
B.4 Data on Head Start 144
B.5 Data on Head Start Launch Dates 145
B.6 The Effect of a Head Start Launch on Head Start Enrollment 148
B.7 Cost-Benefit Analysis of Head Start with the NLSY-79 153
B.8 Additional Estimates 157

BIBLIOGRAPHY 167

LIST OF FIGURES

Figure

Figure		
1.1	Roll-out of STDI pregnancy benefits creates variation over time and across states	44
1.2	Expected labor-market effects of paid maternity leave	45
1.3	Short-run effect on time spent at work in months around childbirth	46
1.4	Evaluating the internal validity of the roll-out of STDI maternity benefits	47
1.5	Effects of paid leave on hourly wages	48
1.6	Effects of paid leave on family income	49
1.7	Effect of STDI maternity benefits on fertility	50
1.8	ITT effect of paid leave enactment on index of educational outcomes	51
1.9	The long-run effects of STDI maternity benefits on children's education	52
2.1	Women's employment, hourly wages, and the advent of STDI maternity benefits	78
2.2	Paid leave leads to decrease in women's relative wages	79
2.3	Impact of enactment of STDI maternity benefits on wage convergence	80
2.4	Paid leave and sorting into professional and management occupations	81
3.1	The Launch of Head Start Between 1965 and 1980	111
3.2	The Expected Pattern of Effects on Adult Outcomes by Age of Child at Head Start's Launch	112
3.3	Funding for Other OEO Programs Relative to the Year Head Start Began	113
3.4	The Effect of Head Start on Adult Human Capital	114
3.5	Visual Representation of Test Statistics Evaluation Pre-Trends and Trend Breaks	115
3.6	The Magnitude of Head Start's Effects on Education Across Studies	116
3.7	The Effect of Head Start on Adult Economic Self-Sufficiency	117
A.1	National share of workers covered by STDI	134
A.2	Estimated share of working women with STDI coverage in 1970	135
A.3	Effect of STDI on time spent at work by month	136
A.4	Correlation of anti-discrimination law passage and state characteristics	137
B.1	1970 School Enrollment by Mother's Education	150

LIST OF TABLES

Table

1.1	Share of new mothers claiming STDI maternity benefits, 1984-1989	39
1.2	Effects of paid maternity leave on hourly wages and employment	40
1.3	Heterogeneity of the effect of paid leave on wages	41
1.4	Long-run effects on child educational outcomes	42
1.5	Estimated effects of educational attainment on potential income	43
2.1	Estimated effects of STDI maternity benefits on the evolution of the gender wage ratio	75
2.2	No effects on hours worked or industry	76
2.3	Effect of STDI on women's relative wages: The role of occupation	77
3.1	The Effect of Head Start on Adult Human Capital	118
3.2	The Effect of Head Start on Adult Human Capital by Sex	119
3.3	The Effect of Head Start on Adult Economic Self-Sufficiency	120
3.4	The Effect of Head Start on Adult Economic Self-Sufficiency by Sex	121
3.5	Heterogeneity in the Effect of Head Start, by Local Programs and Economic Circumstances	122
A.1	Variation in timing and intensity of the expansion of STDI pregnancy benefits	138
A.2	Intent-to-treat effects on working mothers	139
A.3	Effect on working mothers: Complementary evidence from the SIPP	140
A.4	Replication of Gruber (1994): Effect on log wages	141
B.1	Share of Counties and Children under 6 in County with Head Start, 1965-1980	145
B.2	1960 County Characteristics and Head Start's Launch, 1965-1980	146
B.3	Regression Analysis of 1960 County Characteristics and Head Start's Launch, 1965-1980	147
B.4	Age at Launch by Cohort and Year Head Start Launched	148
B.5	Regression-Adjusted Relationship between Head Start and Enrollment	152
B.6	The Effect of Human Capital and Self-Sufficiency on Adult Earnings Potential	156
B.7	The Effect of Head Start on Adult Human Capital by Race	158
B.8	The Effect of Head Start on Adult Human Capital by Race: White Men and Women	159

B.9 The Effect of Head Start on Adult Human Capital by Race: Nonwhite Men and Women . . . 160
B.10 The Effect of Head Start on Adult Self-Sufficiency by Race . . . 161
B.11 The Effect of Head Start on Adult Self-Sufficiency by Race and Race-Sex: White Women and Men . . . 162
B.12 The Effect of Head Start on Adult Self-Sufficiency by Race and Race-Sex: Nonwhite Women and Men . . . 163
B.13 The Effect of Head Start on Incarceration and Mortality . . . 164
B.14 The Effect of Head Start on the Human Capital Index using Different Measures of Access . . . 165
B.15 The Effect of Head Start on the Self-Sufficiency Index using Different Measures of Access . . . 165
B.16 The Effect of Head Start on Completed High School or GED using Different Measures of Access . . . 166
B.17 The Effect of Head Start on Enrolled in College using Different Measures of Accesss . . . 166

LIST OF APPENDICES

Appendix

A. Appendix to The Long-Run Effects of America's First Paid Maternity Leave Policy . . . 124

B. Appendix to The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency . . . 142

ABSTRACT

This dissertation focuses on the interaction between public policy and the U.S. labor market, and its consequences for the economic opportunities available to American women and children. I focus on two public policies designed to enhance opportunities for less advantaged groups: The United States' first large-scale expansion of paid maternity benefits, and the launch of the Head Start preschool program in the 1960s and 1970s. The common thread in these essays is the use of large-scale data and transparent metholodologies to examine the interactions between these policies and individuals' outcomes in the labor market.

The first chapter provides the first evidence of the effect of a U.S. paid maternity leave policy on the long-run outcomes of children. I exploit variation in access to paid leave that was created by long-standing state differences in short-term disability insurance coverage and the staggered enactment of laws that banned discrimination against pregnant workers in the 1960s and 1970s. While the availability of these benefits sparked a substantial expansion of leave-taking by new mothers, it also came with a cost. I find the enactment of paid leave led to shifts in labor supply and demand that decreased wages and family income among women of child-bearing age. In addition, the first generation of children born to mothers with access to maternity leave benefits were 1.9 percent less likely to attend college and 3.1 percent less likely to earn a four-year college degree.

Chapter 2 examines the labor-market consequences of a broad expansion of access to paid maternity benefits. The theoretical implications of maternity leave policies are

ambiguous, with the potential for positive effects that stem from greater attachment to the labor force among mothers but also negative effects that could result from shifts in relative labor demand. I show that the enactment of maternity benefits through STDI slowed the convergence of the gender wage ratio by 31 percent between 1975 and 1985. I also provide evidence that this effect was driven in large part by apparent substitution of men for women into high-profile professional and management positions.

The third chapter, with Martha J. Bailey and Shuqiao Sun, evaluates the long-run effects of Head Start using large-scale, restricted 2000-2013 Census-ACS data linked to date and place of birth in the SSA's Numident file. Using the county-level rollout of Head Start between 1965 and 1980 and state age-eligibility cutoffs for school entry, we find that participation in Head Start is associated with increases in adult human capital and economic self-sufficiency, including a 0.29-year increase in schooling, a 2.1-percent increase in high-school completion, an 8.7-percent increase in college enrollment, and a 19-percent increase in college completion. These estimates imply sizable, long-term returns to investing in large-scale preschool programs.

CHAPTER I

The Long-Run Effects of America's First Paid Maternity Leave Policy

1.1 Introduction

As the role of women in the labor force has grown over the last 50 years, so too has interest in parental leave. Billed as a means to promote child health and help women pursue more continuous, higher-paying careers, nearly every developed nation has adopted policies providing income and job protection to new mothers who take leave as long as one year or more (OECD, 2018). Even in the United States, where maternity leave benefits are allotted far less generously, policymakers and analysts from a wide range of backgrounds have coalesced around the idea that an expansion of paid leave would benefit families and the economy overall (The White House Council of Economic Advisers, 2014; Sholar, 2016).

Despite this growing consensus, little evidence exists on the potential long-run effects of parental leave policies. While a robust body of literature has documented the positive effect that parental leave policies have on the use and length of maternity leave among working women (Han, Ruhm and Waldfogel, 2009; Waldfogel, 1999; Baum, 2003; Berger and Waldfogel, 2004; Baum and Ruhm, 2016; Byker, 2016; Rossin-Slater, Ruhm and Waldfogel, 2013), less is known about the consequences for women's labor-market outcomes or the health and human capital of children. Economists have long understood that mandated parental leave benefits could in theory affect the labor-market prospects of women (Summers, 1989; Gruber, 1994), but recent reviews have concluded that "no obvious consensus on the labor market impact of parental leave rights and benefits emerges from the empirical literature" (Olivetti and Petrongolo, 2017). Furthermore, proponents of parental leave argue that it promotes child health and human capital in the long run by giving mothers and infants more time to bond at a critical period of development (Rossin-Slater, 2018). However, these theoretical effects have proven much more challenging to test empirically, largely because most relevant public policy changes are so recent that the first generations of children exposed to these mandates have not fully reached adulthood.

This paper provides new evidence on parental leave's long-run effects by exploiting a little-studied interaction between U.S. disability policy and anti-discrimination statutes enacted in the 1960s and 1970s. My research design draws on long-standing, cross-state variation in the availability of short-term disability insurance (STDI). These insurance policies, which were originally designed to provide income insurance for temporarily disabled manual laborers, became a source of paid maternity leave benefits when a series of state and federal anti-discrimination laws required them to cover childbirth as a disability. The enactment of these anti-discrimination laws, therefore, expanded paid maternity leave benefits to millions of American women – disproportionately in states where wider STDI coverage gave the policy more "bite."

I use the staggered enactment of these anti-discrimination laws and the pre-existing, cross-state variation in access to STDI to estimate the impact of the enactment of paid leave on a suite of important outcomes, including leave-taking, employment, wages, and the long-run human capital accumulation of children. I use a generalized difference-in-difference empirical strategy that compares outcomes before and after the enactment of STDI maternity benefits, and in states with different pre-existing levels of STDI coverage (Card, 1992). My main analysis relies on data from two sources. First, I use the 1973-1987 Current Population Survey (CPS) May and Multiple Outgoing Rotation Group files to estimate effects on hourly wages, employment, and family income. In addition, I examine the long-run effects on children using confidential microdata from the long-form 2000 Census and the 2001-2016 American Community Survey (ACS) matched to administrative records on the date and place of birth from the Social Security Administration's Numident file. These data represent about one fifth of the U.S. population and provide sample sizes large enough to distinguish potentially small effects.

My research design relies on the assumption that paid leave is the only factor driving the relationship between STDI coverage, anti-discrimination laws, and the outcomes of women and children. I provide several pieces of evidence that suggest this assumption is valid. I find no evidence that the roll-out of paid maternity benefits was correlated with receipt of benefits from the Earned Income Tax Credit (EITC), Food Stamps, and welfare programs. In addition, a balance test proposed by Pei, Pischke and Schwandt (2018) suggests the reform was largely uncorrelated with important changes in demographic characteristics of the population.

Consistent with the theoretical implications of a mandated paid leave benefit, I find evidence that the responses of firms and workers to this new benefit led to a decrease in hourly wages of 4 to 5 percent for women, with no statistically significant changes in women's employment. The effects on wages are highly robust and persistent, but I find no comparable evidence of an effect on men's labor-market outcomes. Moreover, these negative effects on wages and employment combined to generate a decrease in family income that was concentrated among women in the middle of the income distribution.

I also present evidence that this deterioration in women's labor-market conditions imposed costs on the next generation. I find that the children of mothers exposed to STDI maternity benefits achieved worse human capital outcomes in the long run, a result driven by a 1.9 percent decrease in college attendance and a 3.1 percent decrease in the likelihood of earning a 4-year college degree. My estimates of negative effects on women's family income suggests that these results may be driven by a decrease in family resources during children's formative years. The magnitudes of these long-run impacts are consistent with previous estimates of the effect of measures of family resources on child outcomes (Aizer et al., 2016; Stuart, 2018), and they are economically meaningful. For instance, the effect of exposure to STDI maternity leave benefits at birth is large enough to offset roughly one-sixth of the long-run educational benefit enjoyed by Head Start attendees and one-quarter of the benefit accrued to prenatal Medicaid beneficiaries (Brown, Kowalski and Lurie, 2015; Bailey, Sun and Timpe, 2018).

While this paper is the first to report the long-run effects of a maternity leave policy on American children, my results contrast starkly with the benefits enjoyed by Norwegian children born just after an expansion of paid leave in 1977 (Carneiro, Løken and Salvanes, 2015). While the Norwegian and U.S. labor markets have important institutional differences that may contribute to these opposite-signed results, the disparity also highlights a tradeoff inherent in regression discontinuity designs, which often can be used to estimate treatment effects with a high degree of internal validity but may also net out policy-relevant general equilibrium effects. Overall, these estimates suggest that while paid leave policies confer important benefits on working mothers, they may also carry potentially significant costs that should be incorporated in any comprehensive analysis of such policies.

1.2 The creation of America's first paid maternity leave policy

The United States is widely known to be an outlier among developed nations when it comes to parental leave. Roughly 60 percent of workers are eligible for unpaid, job-protected leave through the Family and Medical Leave Act (Klerman, Daley and Pozniak, 2012). In addition, a handful of states have enacted paid family leave programs in the last 15 years. However, no national policy guarantees paid leave for parents who wish to take time away from work before or after the birth of a new child. In fact, while most new parents in Europe and Canada enjoy generous allotments of leave, in 2017 only 15 percent of private-industry workers in the United States report having access to paid family leave (U.S. Bureau of Labor Statistics, 2017).

Less well-known is the fact that many American mothers have access to paid maternity leave through STDI. These policies are required to pay benefits to new mothers by anti-discrimination laws that were enacted nationally in 1979 and even earlier in some U.S. states. The passage of these laws, coupled with pre-existing differences in access to STDI across U.S. states, led to the state-by-state implementation of a paid maternity leave mandate that offers an opportunity to evaluate the long-run effects of such a policy in a U.S. context.

1.2.1 The role of state disability laws and anti-discrimination policy

The U.S. short-term disability insurance industry got its start in the mid-19th century and grew substantially over the next century, driven by the demand for a source of income replacement for temporarily disabled workers (Faulkner, 1940). While coverage varied widely across states and industries, by 1954 the industry covered about 48 percent of workers in most states, with coverage more widespread among unionized workers and large firms (Price, 1986; Levy, 2004). However, coverage is much wider in five states and Puerto Rico, where state law makes access to STDI virtually universal. Rhode Island became the first state to expand access to disability insurance in 1942 when lawmakers created the Cash Sickness Compensation System with the goal of offering wage replacement that nearly all workers could draw on in the case of an illness or injury. California, New Jersey, and New York followed suit in the next few years, while Hawaii and Puerto Rico adopted their own programs in the 1960s (Kamerman, Kahn and Kingston, 1983; Wisensale, 2001). This progression resulted in wide variation across states in access to STDI, with the state-level share covered dependent on the industrial mix in most states but nearly universal for the large fraction of workers in states with STDI guarantees.

This pre-existing variation in STDI coverage became particularly consequential for working women when a series of state and federal laws effectively required them to cover childbirth as a disability. The change came as women's rights groups spoke out against policies around the country that disadvantaged working women, such as insurance policies – including STDI – that excluded coverage of pregnancy. During the 1970s, more than a dozen states enacted policies forbidding discrimination against pregnant workers. While these laws came in a variety of forms – including acts of the legislature in Montana in 1972 and Maryland in 1977, administrative rulings in Kansas in 1975 and Illinois in 1976, and state supreme court decisions such as those in Iowa in 1975 and New York in 1976 – the end result was similar: group STDI plans could no longer exclude childbirth as a covered disability. When Congress approved the Pregnancy Discrimination Act of 1978, the same policy was imposed on the rest of the nation, effectively creating America's first paid maternity leave policy.[1]

1.2.2 Characteristics of STDI maternity coverage

The STDI maternity benefits provided to women were relatively modest by the standards of most OECD countries. They generally covered between one-half and two-thirds of usual weekly wages and lasted between 6 and 10 weeks. While the anti-discrimination laws offered no formal guarantee that a mother's job would be protected, they did require that women on maternity leave receive treatment *equal* to that afforded to others who were absent due to a disability. This formulation could cut both ways: While it afforded "soft" job protection to women at firms that allowed disability leave, it did not preclude employers from uniformly revoking the right to disability leave from all workers.

[1]To assemble evidence on the enactment of state anti-discrimination laws, I rely on several primary and secondary sources, including Congressional testimony, correspondence with state officials, newspaper articles, and published histories of anti-discrimination laws (Gladstone, Williams and Belous, 1985; Kamerman, Kahn and Kingston, 1983; U.S. Senate, 1979; U.S. House of Representatives, 1977). The history of these laws is described further in Appendix A.

In practice, the reform amounted to a large expansion of paid leave at a time when American women received few maternity benefits. Figure 1.1a illustrates the variation in maternity benefit receipt over time that was created by the enactment of anti-discrimination laws in two states with available data, California and New York. The figure plots STDI pregnancy claims as a share of births to residents of each state. With the exception of complications from childbirth, neither state provided STDI benefits to new mothers before pregnancy coverage was extended in 1977. However, the reform led to a sharp increase in benefit receipt, leveling off at roughly 25-30 percent of births or about half of working mothers.[2]

Figure 1.1b shows the differing "bite" that the anti-discrimination laws had across states. The figure displays the share of mothers, by month relative to childbirth, who report receiving STDI benefits in the 1984-1989 panels of the Survey of Income and Program Participation (SIPP). Benefit receipt is much higher in universal-STDI states (solid line) than among women in all other states (dashed line).[3]

Additional context is provided in Table 1.1, which shows that the share of new mothers reporting receipt of STDI benefits around childbirth was 18 percent in universal-STDI states but only 2 percent in other states. This difference is highly statistically significant. The table also shows that claiming was much more common among married women, white women, and women in the middle of the education distribution. These figures suggest the policy was most impactful for middle-class women. The enactment of leave may have been more likely to replicate existing, privately provisioned benefits for highly educated women, while the most disadvantaged groups would have been less likely to work for employers who would agree to an extended absence. In addition, recent survey evidence suggests that women of lower socioeconomic status

[2]Eligibility requirements for STDI benefits are minimal in California and New York, suggesting that the share of eligible mothers can be approximated by the share of New York and California women with a child age 0 who report working for pay in the previous year to the March CPS. This figure hovered between 40 and 50 percent during the late 1970s and early 1980s, suggesting a take-up rate among eligible women of about 50 percent.

[3]Note that these figures imply lower takeup than that implied by the administrative data in Figure 1.1a. This difference is consistent with evidence that receipt of transfer income is significantly underreported in survey data (Meyer, Mok and Sullivan, 2015).

are less likely to be aware of the availability of paid leave (Applebaum and Milkman, 2011).

These differences in the take-up of STDI benefits over time and across states provide *prima facie* evidence of the importance of STDI in the growth of maternity leave among American women. The staggered implementation of anti-discrimination laws at the state and federal levels, combined with long-standing variation in access to STDI, meant that paid maternity leave was expanded differentially across states and time. These policies carried the potential for important effects not only on working mothers, but on children and the entire U.S. workforce.

1.3 Expected effects of paid maternity leave

Discussions of the provision of paid parental leave often focus on its implications for mothers and fathers, the time they spend at home caring for a new child, and their likelihood of returning to a job rather than transitioning to life as a stay-at-home parent. Yet these effects on leave-taking and employment in the short run are only one way that parental leave policies can impact economic and demographic outcomes. Such policies may also have important effects on the employment prospects of the female workforce as a whole by altering women's incentives to work and the hiring and promotion decisions of firms. They may also affect children by changing the mix of time and resources that parents invest in them. Below I discuss the expected effects of STDI-funded maternity leave on each of these groups.

1.3.1 Short-run effects on leave-taking and labor supply

The most immediate effect of the enactment of a maternity leave policy is to alter women's labor-supply decisions in the weeks and months surrounding the birth of a child. New parents face a tradeoff between allocating their time to the firm and home-production tasks related to a child (Klerman and Leibowitz, 1997). In this context, the short-run implications of parental leave policies depend on the presence of two features: wage replacement and job protection. Paid leave benefits reduce the cost of absence from work, leading to greater leave-taking, and may be particularly important in the presence of liquidity constraints (Rossin-Slater, Ruhm and Waldfogel, 2013; Byker, 2016; Bana, Bedard and Rossin-Slater, 2018). Mandated job protection, on the other hand, allows new mothers to take a longer leave than their employers might otherwise be willing to bear. These policies, alone or in concert, should lead unambiguously to an increase in take-up and the length of maternity leave, while the impact on women's attachment to the labor force is less certain. While job protection should increase the share of women returning to the same job after childbirth, paid leave benefits also create an offsetting income effect.[4]

The principal defining feature of STDI-funded maternity leave was its offer of wage replacement in the weeks around childbirth. However, as discussed in Section 1.2.2, these policies also provided "soft" job protection by forbidding employers from treating pregnant women differently than other workers on disability leave. The result is an ambiguous theoretical prediction regarding job retention, but a clear prediction the we should see an increase in maternity leave-taking among working mothers.

1.3.2 Effects on labor demand and supply

In addition to the potential effects on working mothers, the enactment of paid leave mandates may change incentives for firms and the broader set of workers. To illustrate, consider a simple model of a static labor market in a compensating differentials framework. The market includes a unit measure of female workers who make an extensive-margin labor supply decision, $L \in \{0, 1\}$, to maximize a utility function that is increasing in wage income but decreasing in an individual-specific distaste for

[4]A robust body of empirical research on maternity leave has produced evidence largely consistent with these theoretical predictions. New mothers, especially those in Europe, Canada, and other OECD countries, tend to respond to maternity leave mandates by taking more time away from work, while effects on job retention are more difficult to estimate precisely but often positive (Han, Ruhm and Waldfogel, 2009; Waldfogel, 1999; Baum, 2003; Berger and Waldfogel, 2004; Mukhopadhyay, 2012; Baum and Ruhm, 2016; Byker, 2016; Rossin-Slater, Ruhm and Waldfogel, 2013; Bana, Bedard and Rossin-Slater, 2018). Because women who take leave from a job, rather than quitting, retain firm-specific human capital that can translate into higher earnings later on, these results have led a number of commentators to suggest that parental leave policies promote gender equality in the labor market (Waldfogel, 1998).

work, ν_i. This disutility of work, which is distributed in the population according to cumulative distribution function $F(\nu)$, can be interpreted as the cost of maintaining an inflexible work schedule that, for example, limits the amount of time a worker can spend with a newborn child. In that case, we may think of paid leave as a parameter $Z \in [0, 1]$ that moderates the disutility of work by providing greater flexibility. A convenient functional form would be:

$$U(L_i;\ \nu_i) = wL_i - \nu_i L_i Z \tag{1.1}$$

In this simple framework, workers choose to enter the labor force if $w \geq \nu_i Z$; that is, if the market wage is sufficiently high to make up for the inflexibility and other sources of disutility of work. This disutility can be offset if employers take steps to provide workers with more flexibility or reduce other disamenities.

However, efforts to reduce the disamenity of work come at a cost to firms, which must take steps to accommodate extended absences from female workers. Furthermore, the cost of providing flexibility may vary across firms if the absence of a worker is more disruptive in some settings than others. To capture this feature, I model the cost to firm j as a parameter $\delta_j \sim H(\delta)$ that monetizes Z:

$$\pi(L_j) = G(L_j) - wL_j - \delta_j(1 - Z)L_j \tag{1.2}$$

where $G(L_j)$ is a twice-differentiable, concave production function, w is the market wage, and L_j is labor demanded by firm j. Integration of these supply and demand functions leads to the following system of aggregate labor supply and demand that determines equilibrium wages and employment:

$$\textit{Aggregate labor supply}: L^S = \int 1\left\{\nu_i < \frac{w}{Z}\right\} dF(\nu) \tag{1.3}$$

$$\textit{Aggregate labor demand}: L^D = \int L_j^D\left(w + \delta(1 - Z)\right) dH(\delta) \tag{1.4}$$

Equilibrium wages and employment are then determined at equilibrium, where $L^S = L^D$. This simple model replicates the basic insights of Summers (1989) and Gruber (1994). Figure 1.2 provides a graphical representation of the theoretical implications of the introduction of paid leave, which we can think of as an exogenous decrease in Z. The initial equilibrium represented in Figure 1.2a is disrupted by the enactment of paid leave, which makes work relatively attractive to women and shifts the labor-supply curve rightward as shown in Figure 1.2b. In the absence of changes in labor demand, the result would be an expansion of female employment but a drop in wages. However, when we take the response of firms into account, as shown in Figure 1.2c, we see that labor demand will *reinforce* the tendency of wages to fall but *offset* the tendency of employment to rise. Absent any intra-household responses or changes in male labor-market outcomes, which are omitted here for simplicity, these changes could lead to a decrease in income for women even if employment remains unchanged, as shown in Figure 1.2d. An additional prediction is that there will be a sorting effect as the policy elicits a larger demand response among firms where the cost of accommodating maternity leave is higher.

The historical record provides important context when considering the importance of δ_j, the cost of accommodating female workers after the enactment of STDI maternity benefits. After the passage of an anti-discrimination bill in the Maryland legislature in 1977, the state's Chamber of Commerce launched an "urgent" campaign to convince the governor to veto it, arguing that "costs to employers would rise substantially" (Rousmaniere, 1977). In particular, industry representatives objected not only to direct costs of the policy, but also to the cost of replacing workers who would be taking maternity leave rather than returning quickly to their job.[5] Ardie Epranian, a representative of the AVX Corporation, warned members of Congress in a hearing on the Pregnancy Discrimination Act of 1978 that the "real cost is the hidden increase in

[5] Several groups prepared estimates of the cost of expanding STDI maternity benefits while Congress debated the Pregnancy Discrimination Act of 1978, but they varied widely – from a figure of only $130 million from the AFL-CIO to $571 million from the Health Insurance Industry Association. Using data on the annual earnings and family structure of women in the 1976 March Current Population Survey, I estimate that the expected benefits would have amounted to roughly one-half of 1 percent of the annual earnings of the average woman age 18-45.

claims incidence and additional time lost that would be the inevitable consequence... It is rather easy to envision the abuses and extra time lost that can occur." Similarly, a representative of the Electronic Industries Association cited figures from a recent Supreme Court decision, *General Electric Co. v. Gilbert*, that had sided against a woman who sought disability benefits for pregnancy:

> "Other costs associated with this legislation, and I think that some of these have been overlooked, are productivity costs. Employee replacements for women on pregnancy leaves are not as productive as experienced workers. We feel that providing disability benefits will result in longer leaves... It costs money to screen and hire new employees, and as the Gilbert case points out, 40 to 50 percent of females on pregnancy leaves do not return" (U.S. House of Representatives, 1977).

In short, the enactment of paid maternity leave should lead to lower wages for working women but ambiguous effects on employment. This could reflect a reduced willingness to hire women but also a reluctance to promote women within the firm (Thomas, 2018). In addition, to the extent that these changes are driven by labor demand, the negative wage and employment effects may be driven by occupations where leaves of absence are especially disruptive.

Thus far, the literature has produced only limited evidence on the empirical importance of these well-known theoretical implications for the labor market (Olivetti and Petrongolo, 2017; Rossin-Slater, 2018). One limitation, especially in the U.S. context, is related to the fact that the bulk of policy changes have featured complicated eligibility requirements or affected parents in only a handful of states, making inference difficult. Even so, Das and Polachek (2015) and Sarin (2017) study the 2004 expansion of paid family leave in California and find evidence of negative effects on female employment. However, despite the clear theoretical predictions, little evidence has been generated on the effects on women's hourly wages or family income.

A growing body of research has examined the closely related question of whether firms see parental leave policies as costly and respond accordingly. Thomas (2018) finds evidence in the Panel Study of Income Dynamics (PSID) that the job-protected, unpaid leave offered by the Family and Medical Leave Act of 1993 discouraged firms from promoting women to higher-profile positions. However, a different picture emerges in research from Europe, where access to relatively detailed administrative data allows more precise measurements of the effects of generous paid leave policies. Two recent papers using data from Denmark find small effects of maternity leave policies or leave-taking on the success of firms and co-workers (Brenøe et al., 2018; Gallen, 2018). It is not yet clear whether the disparity between findings in the United States and Europe can be attributed to differences in data quality or differences in the setting; given the generous, long-standing social safety net, greater occupational segregation and other features of the labor may make the cost of paid leave less salient to European firms (Blau and Kahn, 2013). Overall, the lack of consensus suggests the debate over the labor-market consequences of parental leave is far from settled.

1.3.3 Effects on children

A final group that may be affected by the enactment of paid maternity leave is the population of children exposed to the policy. Consider the following human capital production function:

$$H = h\left(1 - L_m, w_m, w_f\right) \tag{1.5}$$

where $h(\cdot)$ is a function of the following variables: $1 - L_m$, the mother's time investment mothers make in the child; w_m, the mother's wage; and w_f, the father's wage. The literature on child development suggests H is weakly increasing in each argument (Dahl and Lochner, 2012; Heckman and Mosso, 2014; Agostinelli and Sorrenti, 2018; Almond, Currie and Duque, 2018).

Proponents of paid maternity leave often argue that the effects on children will be positive because the policies increase time investments early in life and, by encouraging greater attachment to the labor force for mothers, increase women's effective wage. The United States' professional association of pediatric physicians has gone so far as to endorse a national paid-leave policy, arguing that "when parents have paid family leave following the birth of a child, mothers breastfeed longer and parents are more likely to take children for immunizations and well-child care... paid family leave can have effects that last throughout life" (American Academy of Pediatrics and Pediatric Policy Council, 2015).

However, the analysis of Section 1.3.2 suggests the paid-leave policy could also lead to a decrease in w_m that could in turn reduce child human capital accumulation. In addition, while the availability of paid leave increases time investments in the child's first months of life, parents may invest *less* time in the long run if the policy encourages greater attachment to the workforce. The ultimate effects on time and resource investments are therefore theoretically ambiguous.

The empirical evidence on parental leave's long-run effects on children offers few hints of the relative importance of these potentially conflicting theoretical forces.[6] The most compelling findings come from Carneiro, Løken and Salvanes (2015), who use a regression discontinuity approach to estimate the effects of an expansion of Norwegian policy from 12 weeks of unpaid leave to 4 months of fully paid leave plus 1 year of unpaid leave. They find substantial effects on children in the long run: a 2 percentage-point decrease in high school dropout rates and a 5 percent increase in wages at age 30. After exploring potential channels, they conclude this effect is the

[6]Studies of short- and medium-run effects on child health or test scores have produced estimates that are generally, but not exclusively, positive. While Ruhm (2000) finds a link between more generous leave policies and lower infant and child death rates in a cross-country analysis, several other papers find no effect on child health and schooling outcomes (Dahl et al., 2016; Baker and Milligan, 2010; Dustmann and Schönberg, 2012; Ahammer, Halla and Schneeweis, 2018). However, Baker and Milligan (2014) find evidence of lower test scores for Canadian children born after an expansion, while Danzer and Lavy (2018) conclude that an Austrian expansion of paid leave led to lower test scores at age 15 among boys with low-educated mothers but benefited boys with mothers who attended post-secondary school. In the United States, Stoddard, Stock and Hogenson (2016) conclude that leave mandates decrease the likelihood of Cesarean delivery, but this effect is reversed if the leave comes with health insurance that would otherwise have been foregone. In addition, two papers associate U.S. expansions of maternity leave with improvements in infant health (Rossin, 2011; Stearns, 2015)

result of increased time spent under the care of the mother, rather than a child-care worker or more distant relative.

While these positive results are striking, several factors limit their generalizability. First, the use of a regression discontinuity design implicitly differences out a number of policy-relevant margins of response for women, such as changes in labor-market conditions, that would be better captured by a difference-in-difference design. In addition, the generous social safety net long present in Norway suggests the labor market may have been better adapted to absorb an expansion of paid leave without a measurable deterioration in wages or employment (Blau and Kahn, 2013). Finally, the Norwegian expansion amounted to an expansion of parental leave allotments for mothers who had already enjoyed more generous benefits than many American workers, even today. Altogether, these considerations suggest reason for caution when using the findings of Carneiro, Løken and Salvanes (2015) to think about long-run, general-equilibrium effects of an expansion of paid leave in the United States.

Estimating such effects in the very long run has been difficult in the U.S. context, largely because most expansions of parental leave were enacted relatively recently – the early 1990s for the unpaid leave granted by the FMLA, and 2004 and later for state paid-leave programs. Another challenge is the availability of data that can link individuals' outcomes as adults to their exposure to the policy as infants, and with sample sizes sufficient to estimate effects with precision. Given the era in which it occurred and the scale at which benefits were expanded, the enactment of STDI-funded maternity leave in the 1960s and 1970s offers a unique opportunity to evaluate these hypotheses in the U.S. context.

1.4 Data and research design

A thorough evaluation of the impact of STDI paid maternity benefits requires data on a wide range of outcomes – including fertility, labor supply, hourly wages, and long-run child outcomes – that are not captured by any single source. I rely on instead on three separate sources of data for my main results.

To document the differential receipt of STDI-funded maternity benefits and the impact on leave-taking and employment in the short run, I construct a sample of women from the 1984-1989 panels of the Survey of Income and Program Participation (SIPP). The SIPP's longitudinal data provides detailed information on labor-market activity and receipt of income from a variety of sources, including STDI. In addition, the 1984 and 1985 panels include retrospective reports on fertility and employment, which I use to construct a month-by-month panel of labor supply for each mother, from 9 months before childbirth to 12 months after.[7] I use these data to examine changes in women's employment and leave-taking around childbirth, as well as their receipt of STDI maternity benefits.

To examine impacts on the broader labor market, I use two sources of data available through the National Bureau of Economic Research (NBER): the Current Population Survey's (CPS) May installment, which provides a continuous measure of hourly wage rates beginning in 1973, and the CPS Multiple Outgoing Rotation Group files, which provide responses to the same hourly wage questions in every month beginning in 1979. Following Lemieux (2006), I use the wage reports of both hourly and salaried workers, dropping imputed values and observations with an hourly wage less than $1 or greater than $100 in 1979 dollars. In addition to hourly wages, I examine effects on employment using the indicator constructed by the Bureau of Labor Statistics, which infers labor-force status from a series of questions about activity in the previous week and other factors. In order to focus on women of child-bearing age and their closest male counterparts, I limit the sample to individuals age 18 to 45. Because earlier years of the CPS do not identify all U.S. states, I consolidate states into 21 groups that can be consistently identified over the course of the sample.

Finally, the estimation of long-run effects requires a source of data that can connect

[7] The survey asks three questions of importance. First, in what year and month did the woman give birth to her first child? Second, did she work during this first pregnancy? And finally, if she did work, when did she stop working before the birth and when, if ever, did she return?

individuals' exposure as children to their economic and demographic outcomes many years later, as well as sample sizes large enough to generate precise estimates of potentially small effects. For this exercise I rely on restricted-use versions of the complete long-form 2000 decennial Census and the 2001-2016 American Community Survey (ACS). These data have been linked to the Social Security Administration's Numident file, which provides a measure of the exact place of birth that has been matched to individuals' county of birth (Stuart, Taylor and Bailey, 2016).[8] To measure outcomes for several years before and after the enactment of paid leave in all states, I restrict the sample to individuals born between 1954 and 1985.[9] I use measures of educational attainment in the Census and ACS to construct four variables of interest: years of schooling and indicators for high school completion, college attendance, and attainment of a four-year college degree. In addition, to increase statistical power, I construct an index of human capital outcomes that consists of the unweighted mean of standardized versions of my measures of educational attainment (Kling, Liebman and Katz, 2007).

Several other public sources of data are used to operationalize and test my research design. These data are described further in the sections that follow.

1.4.1 Research design

The history of STDI maternity benefits suggests a research design that makes use of both the variation in timing of state-level anti-discrimination laws and the differential "bite" of these laws in states with more and less widespread access to STDI. Building

[8]The restricted-use versions of the 2000 Census and 2001-2016 ACS include exact date of birth and state of birth, which is sufficient to infer exposure to the policy. However, the link to the SSA Numident file provides additional flexibility in several ways. First, my preferred specification includes county-of-birth fixed effects, which may improve the precision of my estimates. In addition, observation of county of birth allows me to include specifications that follow previous literature on long-run outcomes by controlling for county-of-birth characteristics and dropping individuals born in large cities such as New York, San Francisco, and Los Angeles (Bailey and Goodman-Bacon, 2015; Hoynes, Schanzenbach and Almond, 2016).

[9]Rhode Island was the first state to pay pregnancy disability benefits, beginning in 1942. Given the advanced age of this cohort in my Census and ACS sample and the difficulty of drawing conclusions from a reform enacted in the middle of World War II, I do not make use of the policy variation in Rhode Island. All other states adopted STDI disability benefits between 1961 (New Jersey) and 1979 (the national Pregnancy Discrimination Act).

on Card (1992), I therefore estimate the following event-study specification:

$$y_{ist} = STDI_{s,1970} \sum_{k \neq -1} \tau_k 1\left\{k = t - T_s^*\right\} + \boldsymbol{\delta_s} + \boldsymbol{\theta_{r(s)t}} + \boldsymbol{X'_{ist}\beta} + \epsilon_{ist} \tag{1.6}$$

where y_{ist} is a measure of women's labor-market outcomes, fertility, or a child's long-run educational attainment and is defined for individual i in state s at time t. This specification includes state fixed effects, δ_s, that control for time-invariant determinants of outcome y_{ist} that may vary across states, as well as a vector of covariates X_{ist} that includes other exogenous determinants of y_{ist}. In my preferred specification, I include fixed effects at the Census-division-by-year level, $\theta_{r(s)t}$, to control nonparametrically for differential trends by region of the country.

The key variable $STDI_{s,1970}$ is designed to capture the variation across states in the share of female workers with access to STDI benefits. Because I do not observe eligibility or receipt of STDI benefits directly, I instead construct a measure of exposure that is not contaminated by firm responses to the anti-discrimination laws. My preferred parameterization of $STDI_{s,1970}$ therefore matches data on female employment by state and industry from the 1970 decennial Census to a tabulation of STDI coverage by three-digit NAICS industry that was prepared by the BLS National Compensation Survey for Autor et al. (2013). This allows me to estimate the share of working women age 18-45 in each state who would have been exposed to STDI maternity benefits:

$$STDI_{s,1970} = \frac{\sum_a \gamma_a \; FemEmp_{as,1970}}{\sum_a FemEmp_{as,1970}} \tag{1.7}$$

where $FemEmp_{as,1970}$ is the number of women age 18-45 employed in industry a in state s in 1970 and γ_a is the national industry-level share of workers with STDI from Autor et al. (2013).[10] In states where STDI is universal, $STDI_{s,1970}$ is assumed to be

[10] An alternative approach would define $STDI_{s,1970}$ as a binary indicator for universal-STDI states. Results using this definition are qualitatively similar and available upon request.

1. This measure of the "bite" of the paid-leave policy thus relies only on the national share of covered workers and the pre-reform industrial mix and disability policy of each state. This specification implicitly assumes that the impact of maternity benefits is proportional to the share of female workers who are covered by STDI. This assumption may be violated if, for example, firms respond to the anti-discrimination policy by dropping STDI coverage altogether in states where it is not required by law. While I cannot rule out such heterogeneous responses altogether, aggregate data suggests such firm responses were not common (see Appendix Figure A.1). In addition, in Section 1.5.1, I explore heterogeneity across universal and non-universal STDI states.

The parameters of interest from equation (1.6), τ_k, can be interpreted as the causal effect of paid leave under the key assumption that the enactment of STDI maternity benefits is the *only* reason that outcome y_{ist} is correlated with my treatment variables. Confounders of this assumption could come in two general forms. First, a trend in y_{ist} over the pre-reform event-time periods would suggest other determinants of the outcome are changing in a way that is correlated with the enactment of paid leave, complicating my estimates of the effect of STDI. Second, a break in unobserved determinants of outcome y_{ist}, if correlated with the enactment of paid leave, would lead me to erroneously attribute the changes in the outcome to STDI maternity benefits.

My flexible event-study specification provides a built-in test of the former assumption. To the extent that confounding pre-trends exist in the data, they would be likely to appear in the form of estimates of τ_k for pre-reform periods that are significantly different from 0. The latter potential confounder is fundamentally untestable. However, I will discuss this assumption further and provide some suggestive evidence of its validity in section 1.4.3.

1.4.2 Take-up of STDI maternity benefits

While the descriptive evidence provided in Figure 1.1 suggests that the enactment of STDI maternity benefits led to an increase in leave-taking among new mothers, this section tests the short-run effects more formally using the regression framework of equation (1.6). To do so, I use the sample of women from the SIPP who respond to the retrospective questions about fertility, limiting the sample to women who gave birth between 1970 and 1984 while between the ages of 18 and 45. Given the relatively small size of the sample, I then restrict the event-time variables of equation (1.6) to a binary indicator for giving birth before or after enactment of STDI maternity benefits. This allows me to estimate a difference-in-difference specification, separately for each month relative to childbirth, to estimate the effect of the policy on the propensity to be with a job and at work.

The results of this exercise are shown in Figure 1.3. For the first two trimesters of pregnancy, the labor supply of first-time mothers changed little as a result of the enactment of STDI benefits, although there is suggestive evidence that the policy led some women to remain in the workforce during the second trimester. Consistent with the structure of most STDI policies, which often covered several weeks before and several weeks after birth, the largest effects come just before and after the month of birth. Women who had access to STDI benefits were roughly 10 percentage points more likely to stay home in the first few months after giving birth. The effect disappears completely by 7 months after childbirth. In short, the policy appears to have achieved paid leave's goal of increasing the time women spend at home with a new child. While I see positive point estimates on labor supply in months 9 through 12, suggesting the potential for increased job retention among new mothers, I cannot rule out effects of meaningful size in either direction.[11]

To get a sense of the impact on time spent at home in the aggregate, we can simply add up coefficients from months -3 through 6, the primary period during which women take maternity leave. This sum amounts to an intent-to-treat effect of -0.56 months, or about 2.4 extra weeks spent at home relative to the counterfactual. However, we can get an estimate of the treatment effect on mothers who received STDI by

[11] In a complementary analysis in Appendix A using decennial Census data that affords larger samples, I find evidence that women with access to STDI maternity benefits were more likely to be employed after childbirth.

scaling these figures by 0.4, my best estimate of the effect of the expansion of STDI maternity benefits on maternity benefit receipt.[12] This exercise suggests that women who received STDI benefits took nearly 6 weeks extra away from work on average. Given that STDI generally provided only between 6 and 10 weeks of wage replacement, this amounts to nearly full take-up of the time allotted by the benefits.

Perhaps unsurprisingly, given the broad nature of the policy and the scarcity of maternity leave allotments at the time, the enactment of STDI maternity benefits compares favorably to more recent expansions of leave policy. For example, in an analysis of California's 2004 paid family leave expansion, Rossin-Slater, Ruhm and Waldfogel (2013) estimate that an extra 6 weeks of paid benefits led to roughly 3 extra weeks of leave for new mothers. The relatively large magnitude of the effect of STDI maternity benefits suggests there may be scope for downstream effects of the policy, as employers may have been more likely to alter their demand for female labor and children may have been more likely to experience a change in their early environment that could have effects in the long run.

1.4.3 Internal validity of the research design

My estimates of the causal effect of paid leave on the outcomes of women and children rely on the assumption that no unobserved determinant of the dependent variable is correlated with the cross-sectional and time variation in access to paid maternity leave. One way to evaluate the validity of this assumption is to estimate equation (1.6) using other indicators that are drivers of women's labor-market conditions or child well-being (Pei, Pischke and Schwandt, 2018). A pre-trend or sharp break in other important determinants of labor-market or child outcomes may be signs that

[12]This figure is calculated as follows: Data from the 1984-1989 panels of the SIPP suggests roughly 18 percent of new mothers receive STDI benefits in universal-STDI states, but only 2 percent in other states (see Table 1.1). While these estimates are known to be downward biased (Meyer, Mok and Sullivan, 2015), if the ratio of these two figures represents the true ratio, then administrative data on STDI receipt among mothers from New York and California suggests 3.3 percent of women in non-STDI states received benefits in the wake of the reform, $\frac{0.02 \times 0.3}{0.18} = 0.033$. The difference in the share of working women covered in the two groups of states is roughly 0.65, which suggests that providing access to paid leave to women results in a change in probably of receiving STDI maternity benefits of $\frac{0.3-0.033}{0.65} \approx 0.4$.

confounding factors are at work.

I focus on two public programs, the Earned Income Tax Credit (EITC) and Food Stamps, which were rolled out during a similar time frame and have been shown to have significant positive effects on women's labor-force participation, children's long-run outcomes, or both (Bastian, 2018; Bastian and Michelmore, 2018; Hoynes, Schanzenbach and Almond, 2016). I construct these variables using state-by-year expenditures from the Bureau of Economic Analysis Regional Income Division and convert them to per-capita terms using the annual population counts from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. In addition, for a measure of public benefit receipt that focuses more directly on the population of interest, I also use the March Current Population Survey (Ruggles et al., 2017) to construct the share of women age 18-45 receiving income from welfare programs and from other government programs, by state group and year, from 1968 to 1984.

The results of this exercise are shown in Figure 1.4. While the EITC's 1975 launch was national, rather than on a state-by-state basis as in the case of paid maternity leave, the program could nevertheless confound my estimates if eligibility or take-up were correlated with the enactment of anti-discrimination laws and the availability of STDI. However, Figure 1.4a suggests little reason for this concern; the trend in per-capita EITC receipt is quite flat and statistically insignificant once I include controls that account for demographic differences across states.

Estimates of the correlation between paid maternity leave and Food Stamps also lead to a null result in Figure 1.4b. After a flat pre-trend, there is a slight increase in Food Stamp benefit per capita after the reform, but the estimates are statistically insignificant.[13] However, an increase following the reform could in fact be partially attributed to maternity leave, if negative effects on female wages led more women to become eligible for the program. If so, this increase in food assistance would be

[13] A joint test of significance of τ_k for the post-reform event-years delivers a p-value of 0.64.

expected to improve children's well-being or at least attenuate any negative effects, given the findings of previous literature on the link between Food Stamps and long-run outcomes (Hoynes, Schanzenbach and Almond, 2016; Bailey et al., 2019).

The results for March CPS measures of the share of women receiving welfare and other government income are also consistent with my identifying assumptions. Figure 1.4c shows no sign of changes in welfare receipt around the reform. Similarly, the trend in Figure 1.4d is flat before the reform and there is no statistically significant evidence of a change afterward.[14]

Overall, the results in this section suggest little reason to think some of the most likely confounders are driving my estimates of effects on female labor-force outcomes and child human capital accumulation.[15]

1.5 Effects on women's employment and wages

The predictions of the stylized model in section 1.3.2 are explored in Table 1.2, which reports estimates from equation (1.6) with τ_k grouped into three-year bins. Column 1 reports the estimated effect on the outcome for which there is a clear prediction, women's log wages. In event years -4 to -2, before STDI maternity benefits were available, I see no effect on wages, consistent with a flat pre-trend. However, wages drop sharply in the first few years after the reform, falling more than 4 percent and

[14]The statistically insignificant jump in the share of women receiving government benefits after the reform may in fact be driven by STDI. In several univeral-STDI states, most notably California, STDI is a state-run program, so beneficiaries may report receiving it under this March CPS category. In fact, the post-reform jump in government income receipt is larger if I restrict the sample to women with a child age 0. This suggests the flat pre-trend and the small increase post-reform are consistent with my identifying assumptions.

[15]In Appendix A I report additional estimates from an exercise that follows Bailey (2006) in testing for systematic relationships between state characteristics and the timing of the roll-out of anti-discrimination laws. I find little evidence that the timing of these state-level laws was correlated with state characteristics as measured in the 1960 Census. The exception is a statistically significant positive correlation between the average education among adult women and the year in which the relevant anti-discrimination law was enacted. While this single statistically significant relationship may well be by chance, given that I perform 21 tests in this exercise, it nevertheless provides a counterpoint to the possibility that early-adopting states were systematically driven by a more educated, empowered female electorate.

remaining at this level even in event years 3 through 5. By contrast, column 2 shows little robust evidence of systematic changes in women's employment.

While the estimated effects on women's wages are strongly statistically significant, there is reason to suspect conventional robust standard errors could be underestimated in settings such as this one, particularly when treatment assignment is clustered (Moulton, 1990; Bertrand, Duflo and Mullainathan, 2004; Kezdi, 2004; Cameron and Miller, 2015; Abadie et al., 2017). One conservative approach to inference in this case is to use a randomization procedure that reassigns treatment assignment at the state level and re-estimates the specification as a test of the null hypothesis that the reform had no effect on wages or employment. In brackets I report p-values from such a procedure using 1,000 replications.[16] Even under this conservative approach, the effect on women's wages remains marginally statistically significant.

Additional detail on the evolution of the effects on wages can be seen in Figure 1.5a, which plots τ_k by event time. My main specification is shown by the navy line with circle markers and confidence intervals. Women's wages were flat in the years leading up to the reform, but this trend broke sharply after the passage of paid leave. The effects remain individually statistically significant even five years after the reform. Figure 1.5a also displays results from several alternative specifications, but the estimated effects change very little, underscoring the robustness of this result.

Do these effects show up in men's labor-market outcomes? The theoretical implications are ambiguous; while we would not expect the enactment of paid maternity leave to have a direct effect on men's labor supply decisions, it could affect intra-household decision-making. In addition, it is possible that labor demand shifted in

[16] Specifically, I calculate these p-values using the following procedure. Consider each realized, state-specific combination of a date of passage of an anti-discrimination law and a share of working women with STDI as a potential value of the treatment. I assign each potential treatment to a randomly selected state (or state group) without replacement. I then estimate equation 1.6, with event-time pooled into three-year bins as in Table 1.2. I repeat this procedure 1,000 times. Under the null hypothesis that the enactment of STDI benefits had no effect on women's wages, I should commonly find that the policy had a large effect even when deliberately mis-assigning treatment in this way. The p-value is calculated as the share of these 1,000 replications that deliver an estimate at least as large in magnitude as my reported estimate.

ways that impact men's wages or employment, with the direction of the effect depending on whether men's labor services are complements or substitutes for those of women. However, the empirical evidence suggests that men saw little or no effect of the policy. The event-study results of Figure 1.5b show no significant effects on the wages of men age 18-45. In line with this visual impression, the results from several specifications in Table 1.2 suggest that the effect on men's wages is small and statistically insignificant.

Given that I observe a significant decrease in wages but little change in female employment or men's labor-market outcomes, a natural question is whether these effects translated to changes in family income. While my sample of May CPS and MORG files do not include measures of family income for my full sample period, the May CPS from 1974-1981 includes a categorical variable corresponding to 13 ranges of family income. I use this variable to construct a series of indicators for family income falling above a given threshold. I then estimate equation (1.6) for each of these thresholds and show the effect at several points in the income distribution.[17]

Figures 1.6a and 1.6b show event-study estimates for two thresholds: The share of families earning more than $1,000 and the share earning more than $7,500, respectively, in nominal terms. I see little effect on family income at the lower threshold. However, there is a clear drop of 2-3 percentage points in the share of families at the higher threshold.

Figure 1.6c shows difference-in-difference estimates at each threshold identified by the May CPS. Consistent with the event-study results, I see little change in family circumstances at the bottom of the income distribution. Families at the top also appear to see little effect. However, families in the middle of the distribution saw statistically significant decreases in the probability of earning above each threshold.[18]

[17] An alternative approach is to use questions from the March CPS. I use the May CPS to ensure my estimates rely on a sample as similar as possible to my earlier wage and employment estimates. However, in practice, I obtain similar results with the March CPS.

[18] Median family income during the middle and late 1970s time frame was between $11,000 and $16,000 in nominal terms (U.S. Census Bureau, 1981).

This suggests that family income was affected most in exactly the families where women were more likely to take up STDI maternity benefits – those in the middle of the skill distribution, as shown in Table 1.1.

1.5.1 Heterogeneity and robustness of labor-market effects

Economic theory suggests the deterioration in women's labor-market outcomes described above comes as a result of an increase in labor supply on the part of women and a decrease in labor demand on the part of firms that worry about the costs of absent workers. In addition, the stylized model of section 1.3 suggests an additional test: To the extent that the cost of accommodating maternity leave varies across firms or occupations, we would expect to see demand shifts of different magnitudes. To test this hypothesis, I follow Hudomiet (2015) and adopt a concept of "adjustment costs" that serves as a proxy for the severity of the disruption a firm would bear due to women taking maternity leave.

To operationalize this concept, I use data from the Multi-City Study of Urban Inequality, which surveyed employers in four U.S. cities between 1992 and 1994 about a range of issues related to hiring and vacancies (Bobo et al., 2008). The survey asked employers how long a new employee would take to become fully productive if hired into a given occupation. I use these data to construct occupation-specific estimates of the adjustment cost and link it to my data from the CPS. I then assign individuals' occupation to above- or below-median adjustment-cost groups, and conduct analyses designed to ask two questions: First, did wages fall more among women in high-adjustment-cost occupations? Second, did working women become more likely to hold a job in a low-adjustment-cost occupation?

The results in Table 1.3 suggest that firm demand did in fact respond more strongly for occupations where absences would be relatively costly. Column 1 tests for a sorting effect by regressing an indicator for working in a high-cost occupation on the specification in equation (1.6). While I find a negative point estimate, it is too imprecise to distinguish from a null effect. However, columns 2 and 3 show that wages did indeed fall further for women in occupations associated with high adjustment costs: I find a 3-percent drop for low-cost occupations but a much larger 8-percent drop in occupations where adjustment costs are above the median. A joint test rejects the null hypothesis that these two estimates are equal, with a p-value of 0.001, suggesting that the enactment of paid leave resulted in disproportionately large wage declines for women whose absence would likely be most costly to the firm.

A second analysis investigates the mechanisms by which paid leave was enacted in different states. As described in Section 1.2, the state-level roll-out of anti-discrimination laws can be divided into two categories: Those in which the law was enacted by a politically representative body such as the legislature, and those in which it was imposed by force that is less responsive to local political pressure, such as the courts or Congress. If the effects of the paid leave were larger in one group of states than the other, it could raise concerns that the results are driven by a selected group of states with fundamentally different political and economic trends.

Columns 4 and 5 of Table 1.3 shows separate wage estimates by category of state anti-discrimination law. In fact, the estimated wage effect in states where the anti-discrimination law was enacted by the legislature or an administrative body is quite similar to the effect where it was imposed from outside, with an estimate of -0.03 in the former and -0.05 in the latter. A test of the equality of these two coefficients delivers a p-value of 0.332, suggesting we cannot reject the null hypothesis that they are equal. These estimates are consistent with the historical narrative, which suggested that the roll-out of anti-discrimination laws was driven more by quirks of the legislative process than systematic differences across states. This bolsters the case that these estimates are picking up the effects of the paid-leave policy rather than other legislation or confounding factors.

Finally, columns 6 and 7 of Table 1.3 provide a check of my research design by splitting the sample by state disability policy. Column 6 provides estimates for universal-STDI states; given the lack of unique identifiers for some small states early in the CPS sample, this group is made up solely of New York and California. Column 7 provides estimates for all other states. The point estimate for universal-STDI states is large and statistically significant, underscoring the binding nature of the policy in those states. However, the result in column 7 makes clear that women in states with lower STDI coverage also saw a decrease in wages; it is smaller, at 3.7 percent, but an F-test cannot reject the null hypothesis that the effect is equal across the two groups of states.

1.5.2 Interpretation of labor-market effects

What drove the deterioration of women's labor-market prospects described in the results of this section? The evidence suggests that firms responded to the enactment of STDI maternity leave by reducing demand for female labor. While positive supply shifts could also lead to lower wages, null or negative changes in female employment suggest that demand was at least as important of a driver. This response by firms is also evident in the larger wage reductions in occupations with high adjustment costs, where we would expect a larger response in labor demand but not supply.

These results are closely related to the effects estimated by Gruber (1994), who evaluated the effect of the Pregnancy Discrimination Act, as well as the corresponding statutes in a handful of states, on employment and wages. The analysis of Gruber (1994) focuses on another consequence of the anti-discrimination laws: The requirement that employer-sponsored health insurance must cover maternity care. This paper exploits similar variation in anti-discrimination policies but also the variation in state STDI coverage.

In Appendix A, I provide evidence that suggests there is reason to reinterpret the findings of Gruber (1994). I exploit the fact that in some states the timing of adoption of STDI maternity benefits was different than the timing of adoption of health insurance benefits. Appendix Table A.4 replicates a key result of Gruber (1994) that suggested the health-insurance mandate led to a 4.3 percent decrease in women's wages. This estimate draws on variation in anti-discrimination laws enacted in three states – New

York, New Jersey, and Illinois. However, when I allow the triple-difference estimate to vary by state, I find that the negative effect is driven by the two states that adopted STDI benefits at the same time they required health insurance policies to cover maternity benefits. In contrast, I find no detectable effect in New Jersey, where the state-run STDI system had been paying benefits for more than a decade before the reform examined in Gruber (1994). This exercise suggests that, while I cannot rule out the possibility that health insurance mandates play some role in my findings, maternity leave was probably the primary driver of the deterioration I observe in women's labor-market outcomes.

1.6 Effects on children

The results outlined in Section 1.5 suggest that women faced significant deterioration in the labor market in the years immediately following enactment of STDI-funded paid maternity leave. There are two channels through which these changes could have affected children. The first would amount to a composition effect if changing labor-market conditions affected women's fertility decisions. The second channel would affect children by altering the investments of time and resources that parents make in their offspring. In the following section I provide evidence that children were impacted primarily by changes in parental investment rather than fertility.

1.6.1 Did paid leave affect fertility patterns?

Given that labor supply decisions are generally thought to be determined jointly with fertility, the changes in women's wages and family income documented in the previous section raise the question of whether women altered their patterns of child-bearing. It is important to understand the effects on fertility because they are of interest in themselves. However, they also play a key role in the interpretation of the long-run effects on children described in the next section. If paid leave altered the size of cohorts born in the wake of the reform or changed the composition of the group of women bearing children, this could lead to a selection effect that drives changes in average outcomes for the group years later.

To test for changes in fertility, I assemble a state-by-month panel using birth records from the 1974-1984 Natality Detail Files, available through ICPSR, and population counts from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. I examine effects on the fertility rate, birthweight, and mother's race, since changes in any of these characteristics could be signs that women altered their child-bearing patterns. In addition, because the quality of education data in the Natality Detail Files is low during this time frame, I use a sample of children from the 1980 decennial Census to test for changes in the composition of women giving birth, as proxied by years of education.

Figure 1.7a shows estimates of the effect on the fertility rate from (1.6). Despite the visual impression of a dip in fertility rates after the enactment of paid leave, formal tests of the significance of these effects suggest we cannot distinguish them from 0.[19] Nor do I see compelling evidence of a change in birthweight, a common marker of child health, which is displayed in Figure 1.7b. Finally, my estimates for a change in the composition of women giving birth is also null, with no apparent effect on the mother's race or years of education in figures 1.7c and 1.7d.

These estimates suggest there is a little scope for fertility changes to drive long-run effects on children's long-run outcomes. If anything, the slight but statistically insignificant rise in average weight at birth and drop in share nonwhite suggests positive selection of the cohorts born immediately after the reform. As I will show in the next section, the ultimate effect on the long-run outcomes of these first exposed cohorts suggests any positive selection that may have existed was offset by other factors.

[19] A test of the joint significance of the coefficients on event time 0 through 2 in Figure 1.7a results in a p-value of 0.71.

1.6.2 Effects on children in the long run

My estimates of the intent-to-treat effect of STDI maternity benefits on the index of children's long-run education outcomes, constructed as the unweighted men of standardized versions of my primary outcomes, is shown graphically in Figure 1.8. The red line with triangle markers shows results from a specification that includes only county of birth fixed effects, fixed effects by birth and survey years to control nonparametrically for age and cohort effects, and a month of birth fixed effect to control for the seasonality in socioeconomic status of new births (Buckles and Hungerman, 2013). The green line with circle markers shows results from a specification that adds fixed effects at the year of birth by Census division level, which accounts nonparametrically for trends that vary across regions of the United States. Finally, my preferred specification, the blue line with no markers, follows the literature on long-run effects by adding predetermined characteristics of the county of birth interacted with a linear trend in year of birth (Almond, Hoynes and Schanzenbach, 2011; Bailey and Goodman-Bacon, 2015).[20]

The stability of my estimates across these three specifications underscores the robustness of the result: A drop in the human capital accumulation of the first generation born to mothers who were eligible for STDI maternity benefits. If these effects are driven by reductions in female wages and family income, we would expect spillover effects on older children who were born before the reform but are nevertheless exposed to some extent to the decrease in family resources. The slight negative slope of the pre-reform coefficients is consistent with this explanation; however, a joint test of the pre-reform coefficients in Figure 1.8, with a p-value of 0.43, fails to reject the null hypothesis that there is no pre-trend in educational outcomes. The flexible event-study specification also allows us to distinguish the dynamic effects of the policy after the reform. After a sharp break downward at event-time 0, the negative effects continue

[20]These characteristics include county-level measures from the 1960 Census: Share of population living in an urban area, share in a rural area, share under 5 years of age, share over 65 years of age, share nonwhite, share with 12 years of education or more, share with less than $3,000 in annual income, and share with greater than $10,000 in annual income. Each of these characteristics is interacted with year of birth.

to grow in magnitude as more women take up STDI benefits (see Figure 1.1a) and the labor-market conditions experienced over the lifetime of the average mother continue to deteriorate.

These results are summarized in row 1 of Table 1.4. Column 2 shows the result of a simple difference-in-difference estimate estimate of the intent-to-treat effect on the index of educational outcomes; children who were exposed to the policy saw a decrease of 1.9 percent of a standard deviation. Column 3 shows the F-statistic and p-value from a test of the null hypothesis that the pre-reform event-time coefficients are jointly equal to 0. The p-value of 0.43 provides little cause to reject the null hypothesis that there is no confounding pre-trend in the outcome variable.

While these negative effects on the index of education outcomes suggest children exposed to maternity leave saw a deterioration in their long-run outcomes, these results tell us little about the magnitude of the impact. For additional context, Table 1.4 also includes estimates for each of the four components of the index: years of education, high school completion, college attendance, and 4-year college completion. Column 2 suggests that children exposed to the paid-leave policy achieved 0.05 fewer years of education, or a decrease of two-fifths of 1 percent. The remaining results suggest that this decrease in educational attainment was concentrated at the upper end of the distribution: while the effect on high school completion is statistically indistinguishable from 0, college attendance fell by 1.9 percent among exposed children, and college completion fell by 3.1 percent. Figure 1.9 shows the results for college attendance and completion graphically, and they are qualitatively similar to the effects on the educational index, with downward break in the trend that levels off only after four to five years.

What could explain these sizable decreases in educational attainment for children exposed to STDI maternity benefits? One possible mechanism is a decrease in investment of resources in the first generation of children born after enactment of paid leave. Using data from the March CPS accessed via IPUMS (Ruggles et al., 2017), I calculate that the effects on family income reported in Section 1.5 amounted to a decrease of about 2 percent in family resources. If we assume this effect was persistent, and that the negative effects are driven solely by this change in family income, it suggests that a maternity-leave driven decrease in family income of 10 percent leads to a reduction in years of schooling of roughly one quarter of a year, or nearly 2 percent, and a nearly 5 percentage-point decrease in four-year college degree attainment.

These large effects are comparable to estimates of long-run impacts from other settings. For example, Stuart (2018) finds a 3 percentage-point decrease in college degree attainment for every 10 percent decrease in earnings per capita driven by the double-dip recession of the early 1980s. Similarly, in a study of an early welfare program, Aizer et al. (2016) conclude that an early welfare program raised family income during childhood by 20-30 percent and schooling by 4.3 percent of the control-group mean; an extrapolation of my results would suggest that a drop of income of similar magnitude would reduce years of schooling by a comparable 4 to 6 percent. While the settings examined by these papers are quite different from that of the expansion of STDI maternity benefits, the similar magnitudes of the effects provide assurance that the deterioration in child educational outcomes could reasonably be driven by the unintended decrease in family income.

Another way to place my estimates in context is to compare them to estimates of the long-run effects of other policies designed to improve children's long-run outcomes. For example, Bailey, Sun and Timpe (2018) examine the roll-out of Head Start and find an intent-to-treat effect of 0.29 extra years of schooling for children who attended fully implemented programs, or 0.043 years for all children exposed to the launch of a local Head Start center. An expansion of Medicaid coverage for pregnant women increased their children's high-school completion rates by nearly 4 percentage points, with suggestive evidence of effects of a similar magnitude on college attendance (Miller and Wherry, 2017). Similarly, Brown, Kowalski and Lurie (2015) find that a year of Medicaid enrollment in childhood raises the probability of enrolling in college by age

20 by 0.55 percentage points.[21] My estimates suggest that the magnitude of the effect of the enactment of paid leave was roughly one-sixth the size of the long-run educational-attainment benefit received by Head Start attendees, one-quarter of the college-attendance benefit enjoyed by beneficiaries of Medicaid while in utero, and the equivalent of about two years of Medicaid coverage in childhood. Overall, these results suggest the enactment of paid maternity benefits sparked a series of changes in the labor market and ultimately children's outcomes, with a magnitude similar to that of some the United States' most highly touted public programs, but in the opposite direction.

1.6.3 Reconciling long-run estimates with previous literature

My findings may be surprising in light of theoretical literature that emphasizes the importance of mother-child bonding time during critical periods of life, as well as empirical evidence that has suggested maternity leave policies improve infant health. While I do not find positive impacts on infant health, my results do not necessarily contradict the hypothesis that infants benefit from the increased bonding time and reduced stress conferred by paid maternity benefits. Rather, they suggest that to the extent such benefits exist, they are at risk of being attenuated or even reversed by unintended consequences of maternity leave policies, such as a deterioration of labor-market conditions that leaves families with fewer resources to invest in children during their formative years.

My findings are also at odds with the results of Carneiro, Løken and Salvanes (2015), who study a 1977 maternity leave expansion in Norway and find large positive effects on children in the long run. Two key differences may help reconcile these disparate findings.

The first key difference is related to research design and its implications for the in-

[21] Brown, Kowalski and Lurie (2015) find an effect on college attendance of 0.24 percentage points for men and 0.4 percentage points for women. The simple average of these two figures, divided by an estimate of 0.58 years of enrollment per year of eligibility, delivers an estimate of 0.55 percentage points.

terpretation of the estimates. Given the sharp policy change and rich data available, Carneiro, Løken and Salvanes (2015) use a regression discontinuity design that effective compares children born under a more generous policy regime to those born just a few days earlier. Such a design approximates an experiment in which paid maternity benefits are randomly assigned to expectant mothers, ruling out the possibility of general-equilibrium effects such as changes in the labor market that could differentially affect the treatment and control groups. The policy experiment in this paper, on the other hand, approximates a more general – and, arguably, more policy-relevant – experiment in which women and firms are allowed to respond across all possible margins to the introduction of paid leave benefits. In short, while Carneiro, Løken and Salvanes (2015) demonstrate compellingly that increased mother-child bonding time can lead to valuable improvements in long-run human capital accumulation, my results demonstrate that such effects can also be reversed by the unintended consequences of mandated maternity benefits.

The second key difference is related to the context of the two studies. The natural experiment examined by Carneiro, Løken and Salvanes (2015) took place in a country with a long-standing, relatively generous social safety net, including subsidies for relatively high-quality child care. Labor demand responses may be muted in countries where a higher degree of occupational segregation makes maternity leave less costly from the perspective of the firm (Blau and Kahn, 2013). In fact, findings from the nascent literature examining firm responses to maternity leave mandates suggests just such a dynamic, with firms in the United States displaying elastic demand for female labor while European firms respond less dramatically to parental leave policy (Thomas, 2018; Brenøe et al., 2018; Gallen, 2018).

1.7 Conclusion

The robust body of literature on the effects of family leave policies has demonstrated clearly that parents, and especially mothers, greatly value the opportunity to take an extended absence after the birth of a child without surrendering a job match or the stream of income that comes with it. However, the recent nature of U.S. parental-leave policies has made it difficult to evaluate effects that may take years or even decades to materialize.

This paper provides the first estimates of these long-run effects in the United States by constructing a history of the country's first expansion of paid maternity leave. While the policy greatly expanded the availability of maternity benefits and increased the amount of time working mothers spent on leave with their newborn children, it also generated a response from employers who feared that this new workplace flexibility would make women more costly to employ. I find evidence that the interaction between women's responses to the policy and shifts in firms' relative labor demand led to a 4-percent decrease in women's wages and a deterioration in the incomes of middle-class families. Furthermore, these effects persisted into the next generation, reducing children's educational attainment by 0.05 years and decreasing their probability of attending or graduating from college by 1.9 and 3.1 percent, respectively.

The finding of negative wage and family income effects, paired with a long-run decrease in children's educational attainment, suggests maternity leave policies may not necessarily achieve the goal of promoting gender equity and improving the welfare of the next generation. On the contrary, my results suggest that parental leave may bestow benefits on parents and their children in the short run while accruing significant costs in the long run.

How do we weigh these long-run costs against the short-term benefits of an expansion of paid maternity leave? In the literature on long-run effects of childhood interventions, one common way to quantify these costs is to generate an internal rate of return on the resources invested in the child. This exercise provides a way to scale the benefits – or, in this case, the costs – by relating their discounted future value to the initial amount invested.

To calculate the internal rate of return on STDI maternity benefits, I follow previous literature and first convert my estimates of the long-run effects on children's education to effects on potential earnings (Neal and Johnson, 1996; Deming, 2009; Bailey, Sun and Timpe, 2018). While realized earnings may be affected in subtle ways by changes in selection into the workforce, the impact on potential earnings can provide a sense of the opportunities gained or lost as a result of the treatment. Using a sample from the National Longitudinal Survey of Youth (NLSY) 1979 cohort, I regress log earnings on educational attainment and demographic covariates.[22] The use of the NLSY allows me to include AFQT scores, a proxy for ability, in the specification to alleviate concern about omitted variables bias. I then convert these estimates to the present value of lost potential earnings between age 25 and 54. If we scale this figure by the STDI maternity benefit take-up rate and compare it to the average benefit, we get a sense of the internal rate of return of the initial investment to the average child. Note that this calculation is inherently conservative because it abstracts from the cost of raising the funds and the immediate costs of the decrease in wages and family income that resulted from the reform.

My estimates from the log-earnings equation are shown in Panel A of Table 1.5. To facilitate comparisons to my long-run results, my preferred specification includes a linear term in years of schooling plus dummy variables to capture the effects of completing high school, attending college, and graduating from a four-year college. Column 1 includes only controls for education, age, race, and survey year. The effect of accounting for a proxy for underlying ability can be seen clearly in column 2, where I add a quadratic in AFQT to the specification and the coefficients on education fall considerably. For robustness, column 3 shows that I obtain similar estimates from a more standard specification where an indicator for college attendance is omitted. Finally, columns 4 and 5 break down the sample by gender, showing that returns to high school are higher for women but that college attendance and graduation are particularly profitable for men.

Panel B summarizes the implied effects on potential earnings between ages 25 and

[22] To match the individuals in my sample from the 2000 Census and 2001-2016 ACS, I use individuals only over the age of 25. I also drop individuals older than 54 to avoid concerns about retirement. Earnings has been converted to 2012 dollars using the CPI-U.

54. Data from the state of New York suggests that the average mother who benefited from STDI between 1978 and 1985 received $3,129 in 2012 dollars.[23] In my preferred specification, the education effects suggest a decrease in potential earnings of about one-half of 1 percent per year. At birth, assuming a 5 percent discount rate, this equates to a cost of roughly $532 in 2012 dollars. However, if we scale it using a conservative estimate of a 25 percent STDI take-up rate, it becomes clear that the cost per treated child is much higher: more than $2,000 in discounted earnings, or an internal rate of return of -68 percent. Although the effects on education are comparable by gender, the differences in the return to college and potential earnings make the implied effect on men much larger: An 80 percent IRR for men relative to 58 percent for women.

This simple calculation does not account for the potential social costs of paid leave, including the cost of raising funds and the decreases in family income that occurred immediately in the wake of the policy. Even so, it suggests that maternity benefits come with significant long-run costs. These very large and negative internal rates of return are driven largely by the fact that the cost of paid leave is not limited to, and perhaps not even driven by, the direct cost of the benefit. Rather, the provision of these relatively modest benefits triggered large responses in the labor market because of the cost of disruptions for firms, whether real or perceived. An additional consideration is that, while my data does not allow me to observe variables about the family characteristics of the children who were affected, data from the SIPP suggests women who made use of STDI maternity benefits were relatively advantaged. This observation raises questions about the distributional consequences of the policy. Overall, the enactment of STDI maternity benefits suggests that future paid-leave policies must take into account the potential that they could alter labor-market opportunities for women and, by extension, the well-being of future generations.

[23] Data on STDI pregnancy benefits paid nationally is generally not available. I use New York's figures because the benefit amounts were relatively modest (50 percent of weekly earnings up to a cap) and the state Workers Compensation Board provided reports that include claims, average length, and total payments by year.

Table 1.1: Share of new mothers claiming STDI maternity benefits, 1984-1989

	(1) Universal STDI states	(2) All other states	(3) P-value: Test of difference
All mothers	0.18	0.02	0.000
	(0.39)	(0.15)	
Age 18-29	0.19	0.02	0.000
	(0.40)	(0.14)	
Age 30-45	0.16	0.03	0.001
	(0.37)	(0.16)	
Married	0.21	0.02	0.000
	(0.41)	(0.16)	
Unmarried	0.08	0.01	0.004
	(0.27)	(0.10)	
Nonwhite	0.13	0.01	0.001
	(0.33)	(0.09)	
White	0.20	0.02	0.000
	(0.40)	(0.16)	
HS dropout	0.07	0.01	0.002
	(0.26)	(0.07)	
HS grad & some college	0.22	0.02	0.000
	(0.42)	(0.16)	
Four-year college graduate	0.18	0.03	0.009
	(0.39)	(0.16)	
Observations	1,265	4,486	

Notes: Data comes from sample of women age 18-45 who give birth during the 1984-1989 panels of the Survey of Income and Program Participation. Column 1 shows share receiving STDI maternity benefits during the third trimester, the month of birth, or the three months after birth in universal-STDI states of California, New York, New Jersey, Hawaii, and Rhode Island. Column 2 shows share receiving benefits in all other states. Standard deviations are in parentheses. Column 3 shows p-value from test of null hypothesis of no difference in share receiving benefits across the two groups of states.

Table 1.2: Effects of paid maternity leave on hourly wages and employment

	Women		Men	
	(1)	(2)	(3)	(4)
	Log wages	Employed	Log wages	Employed
Event years -4 to -2	0.000717	0.0158	0.0126	-0.0101*
	(0.0100)	(0.0110)	(0.0117)	(0.00560)
	[1.00]	[0.27]	[0.64]	[0.62]
Event years 0 to 2	-0.0436***	-0.00110	-0.00623	-0.0114
	(0.0128)	(0.00473)	(0.0137)	(0.00665)
	[0.09]	[0.89]	[0.81]	[0.39]
Event years 3 to 5	-0.0432**	0.0134*	-0.00247	-0.0120
	(0.0193)	(0.00766)	(0.0153)	(0.00915)
	[0.15]	[0.28]	[0.95]	[0.43]
Observations	584,761	1,063,681	673,816	973,623
R-squared	0.271	0.063	0.357	0.114
Control mean	4.22	0.630	5.93	0.847

Notes: Coefficients displayed are estimates of τ_k from equation (1.6) with event time pooled into three-year bins. Standard errors in parentheses are clustered by state group. Figure in brackets is p-value from a randomization inference procedure based on 1,000 draws of state-level STDI coverage and anti-discrimination law enactment date. All specifications include a quadratic in age interacted with indicators for Hispanic ethnicity and nonwhite race, years of education, indicators for completing high school and four-year college, and fixed effects for year-by-month, state-group, and Census-division-by-year. Specification also includes linear trend in survey year interacted with the following state-level characteristics from the 1970 decennial Census via IPUMS (Ruggles et al., 2017): share black, average years of education among women, share with high school degree, share with college degree, number of children born to women, and share in poverty. Sample includes men and women age 18-45 from the 1973-1987 May and Merged Outgoing Rotation Group CPS files. Individuals with imputed values have been dropped, as have wage observations below $1 or above $100 in 1979 dollars. Wages are converted to 1979 dollars using the CPI. Regressions are weighted using CPS earnings weights.

Table 1.3: Heterogeneity of the effect of paid leave on wages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	High-cost job	Log wages		Log wages		Log wages	
	All employed women	Low-cost occupations	High-cost occupations	Legislative reform	Congress or courts	Universal STDI	Non-universal
Event years -4 to -2	-0.0132*	0.00180	0.00638	-0.0136**	0.00745	-0.00597	0.000799
	(0.00734)	(0.0106)	(0.0141)	(0.00622)	(0.0153)	(0.00621)	(0.0168)
Event years 0-2	-0.0172	-0.0309***	-0.0836***	-0.0301**	-0.0529***	-0.0641***	-0.0370*
	(0.0121)	(0.0107)	(0.0116)	(0.0119)	(0.0168)	(0.0193)	(0.0204)
Observations	791,386	582,514		584,761		584,761	
R-squared	0.182	0.302		0.271		0.271	
F-statistic: Equal effects		14.70		0.988		0.804	
P-value: Equal effects		0.00104		0.332		0.381	

Notes: Coefficients in column 1 are estimates of τ_k from equation (1.6) with event time pooled into three-year bins, and $STDI_{s,1970}$ calculated using equation (1.7). In columns 2-7, a separate τ_k is estimated for each group specified. Dependent variable in column 1 is a dependent variable indicating employment in an occupation in which the time required to become fully productive is above the median, as measured in the Multi-City Study of Urban Inequality (Bobo et al., 2008). In columns 2 and 3, effect on log wages is estimated separately for for low- and high-cost occupations. In columns 4 and 5, effect on log wages is estimated separately for states where paid leave was enacted due to an act of the Legislature or executive branch (column 4) or due to a state supreme court decision or act of U.S. Congress (column 5). Sample includes men and women age 18-45 from the 1973-1978 May CPS and 1979-1987 Merged Outgoing Rotation Group CPS files. Standard errors in parentheses clustered at state group level. Wages are converted to 1979 dollars using the CPI.

Table 1.4: Long-run effects on child educational outcomes

	(1) Sample mean	(2) Intent-to-treat effect	(3) Test for pre-trend	(4) Percent change
HC index		-0.0186***	0.97	
		(0.004)	[0.43]	
Years of schooling	13.7	-0.0538***	1.71	-0.4%
		(0.009)	[0.16]	
High school graduate	0.93	-0.00118	1.05	-0.1%
		(0.001)	[0.39]	
Some college	0.66	-0.0125***	0.4	-1.9%
		(0.003)	[0.81]	
College graduate	0.32	-0.00998***	1.44	-3.1%
		(0.002)	[0.23]	

Notes: Coefficients displayed in column 2 are estimated intent-to-treat effects of exposure to paid maternity benefits on children in the long run. Sample includes individuals in the 2000 long-form decennial Census and 2001-2016 American Community Survey linked to the Social Security Administration's Numident file, born in the United States between 1954 and 1985 and age 25 or older when surveyed. Column 3 shows F-statistic and p-value from a test of the null hypothesis that the pre-reform coefficients are jointly equal to 0. Column 4 shows estimate as a percent change relative to sample mean. Standard errors in parentheses clustered at state of birth level.

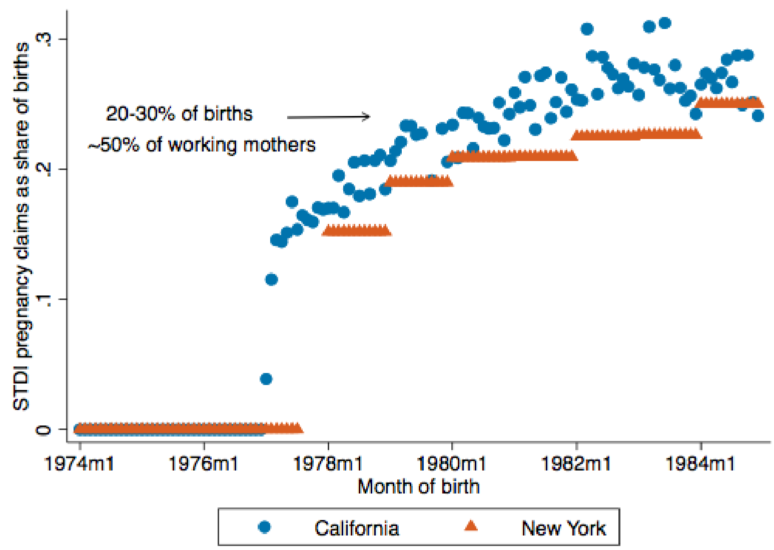
Table 1.5: Estimated effects of educational attainment on potential income

	(1) Log earnings	(2) Log earning	(3) Log earnings	(4) Log earnings	(5) Log earnings
Panel A: Log earnings estimates					
Years of education	0.0175*	0.00328	0.00781	0.0252*	-0.0185
	(0.00944)	(0.00950)	(0.00759)	(0.0138)	(0.0126)
High school degree	0.355***	0.227***	0.224***	0.340***	0.231***
	(0.0395)	(0.0394)	(0.0417)	(0.0675)	(0.0505)
Some college	0.177***	0.106***		0.0960**	0.126***
	(0.0274)	(0.0273)		(0.0396)	(0.0376)
College degree	0.314***	0.288***	0.289***	0.243***	0.317***
	(0.0367)	(0.0364)	(0.0373)	(0.0536)	(0.0489)
AFQT	No	Yes	Yes	Yes	Yes
Sample	All	All	All	Women	Men
Observations	129,536	129,536	129,536	62,838	66,698
R-squared	0.204	0.222	0.183	0.138	0.199
Mean	26,370	26,370	26,370	20,848	33,190
Panel B: Discounted value of change in potential earnings					
Annual change	-176	-122	-94	-104	-143
Total discounted value	-768	-532	-409	-452	-625
Total per treated child	-3072	-2129	-1637	-1807	-2501
Internal rate of return	-98%	-68%	-52%	-58%	-80%

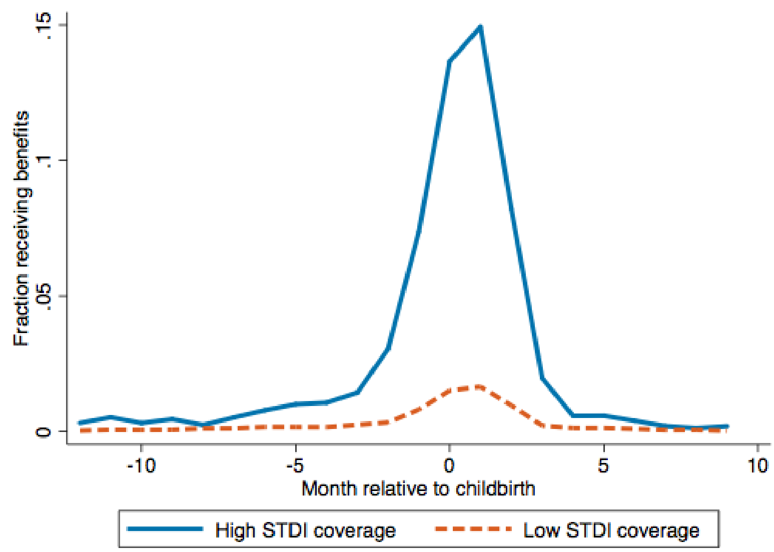
Notes: Data includes individuals from National Longitudinal Survey of Youth 1979 cohort, age 25 and older. In addition to education variables shown, specifications include survey year, quadratic in age interacted with race and gender, and a quartic in AFQT score. AFQT score has been standardized within the sample by year of birth. Standard errors are clustered by individual to adjust for within-person correlation in error term over time. Discounted value of lost potential earnings assumes fixed discount rate of 5 percent. Total per treated child assumes take-up rate of STDI benefits of 25 percent. Internal rate of return is constructed using an estimated average STDI maternity benefit of $3,129 in 2012 dollars. All figures are expressed in 2012 dollars, adjusted using the CPI.

Figure 1.1: Roll-out of STDI pregnancy benefits creates variation over time and across states

(a) Launch of STDI benefits in two high-coverage states



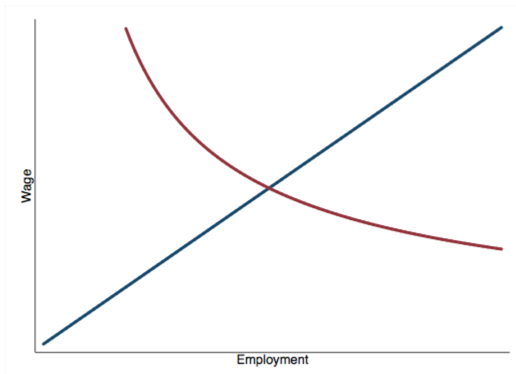
(b) Take-up of STDI benefits wider in high-coverage states



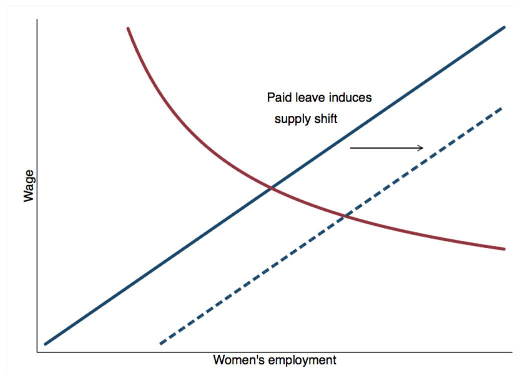
Notes: STDI maternity benefits were enacted in January 1977 in California and August 1977 in New York. Data in Figure 1.1a is constructed by dividing the number of STDI pregnancy claims by month or year in California and New York by the number of births to residents of those states. STDI pregnancy claims provided by California Employment Development Department and New York Workers Compensation Board. Birth records come from Natality Detail Files (National Center for Health Statistics, 2015). Data in Figure 1.1b comes from sample of women age 18-45 who gave birth during the 1984-1989 panels of the Survey of Income and Program Participation. Solid line shows share of women receiving STDI maternity benefits, by month relative to childbirth, in the universal-STDI states of California, New York, New Jersey, Hawaii, and Rhode Island. Dashed line shows share receiving benefits by month in all other states.

Figure 1.2: Expected labor-market effects of paid maternity leave

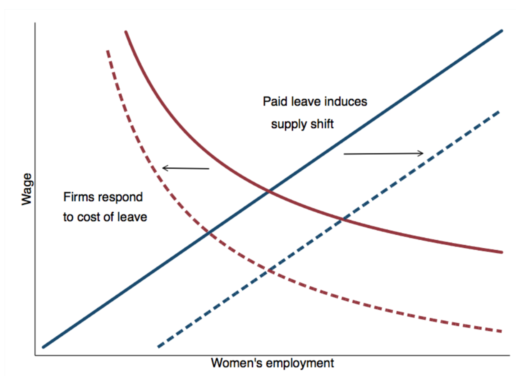
(a) Labor supply and demand



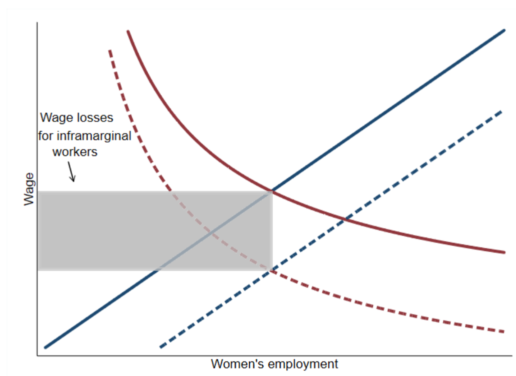
(b) Labor supply shifts



(c) Labor demand response

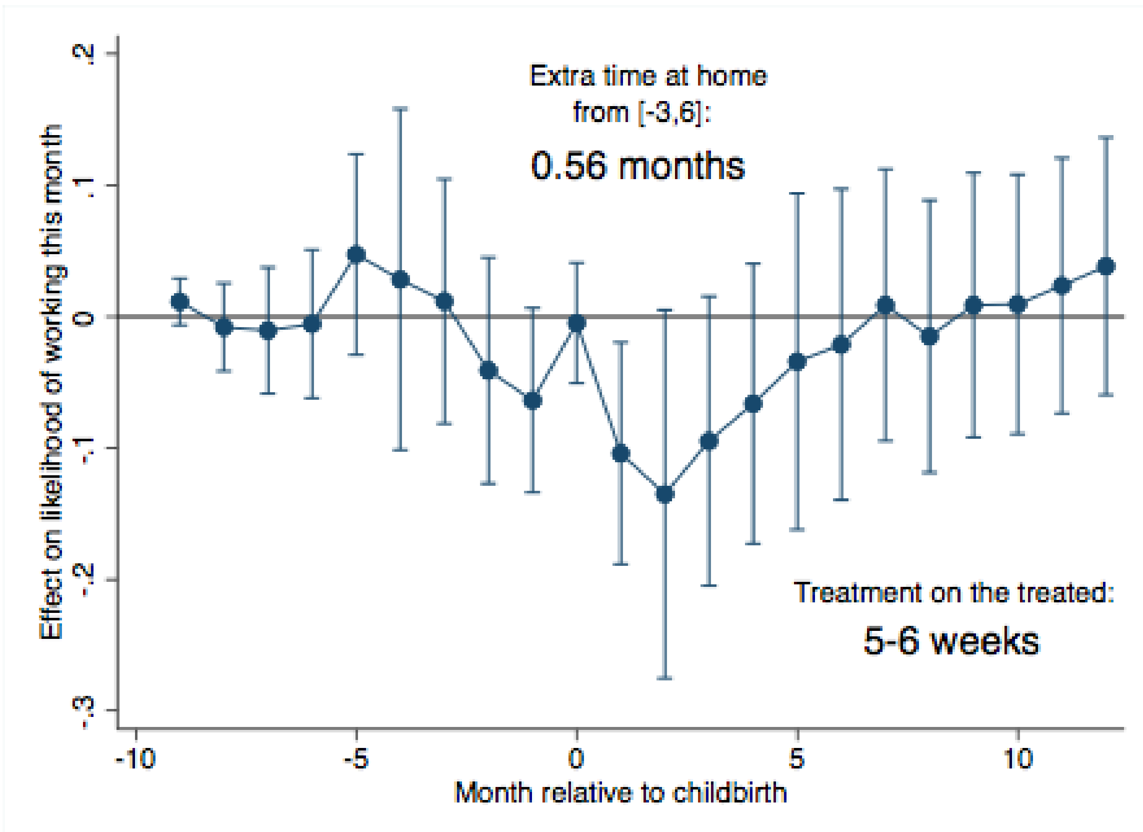


(d) Income loss for inframarginal women



Notes: Figure shows graphical representation of stylized labor-market model outlined in Section 3. Panel 1.2a shows initial labor-market equilibrium. Panel 1.2b shows response of women to enactment of benefit that reduces disutility of work. In Panel 1.2c, firms respond to the cost of providing the benefit. Panel 1.2d shows the impact of wages lost among inframarginal workers who are impacted by the change in the equilibrium wage but would have remained in the labor force in the absence of paid leave.

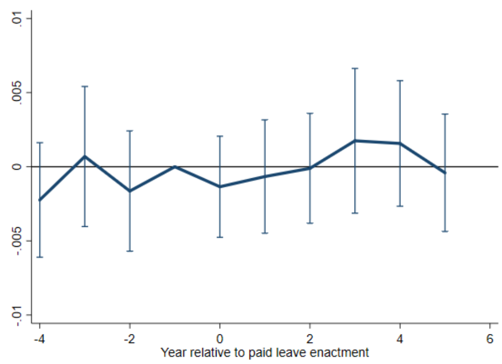
Figure 1.3: Short-run effect on time spent at work in months around childbirth



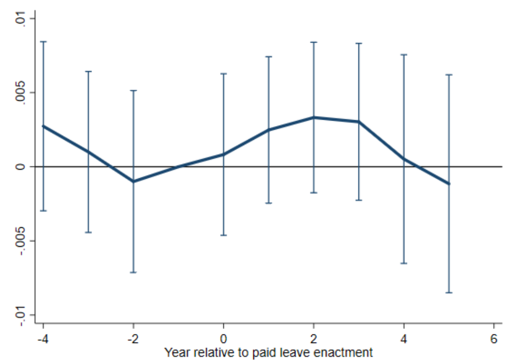
Notes: Data includes women from the retrospective fertility module in the 1984 and 1985 SIPP. Sample is limited to women whose first child was born between 1970 and 1984 while between the ages of 18 and 45. Women are asked about labor supply by month only if they worked during their first pregnancy. Figure shows intent-to-treat estimates of STDI exposure on time spent at work by month relative to childbirth, using a version of equation (1.6) that restricts event time to dummies indicating birth before or after the reform. Standard errors in Panel B are clustered at the state-group level.

Figure 1.4: Evaluating the internal validity of the roll-out of STDI maternity benefits

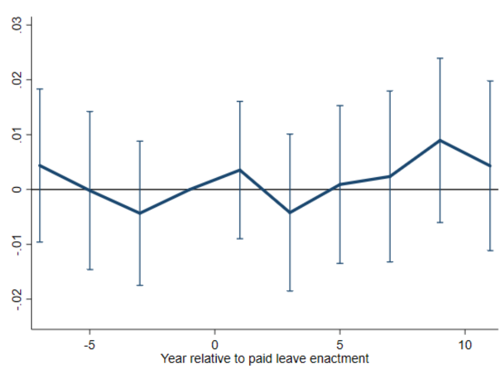
(a) EITC per capita (tens of dollars)



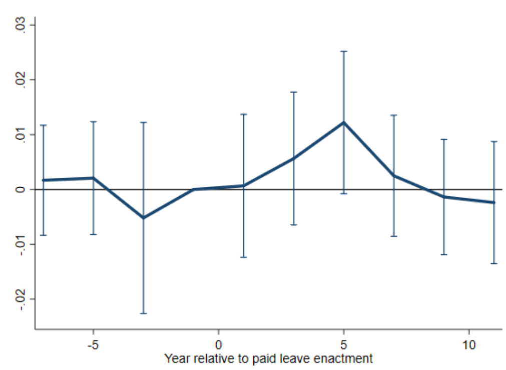
(b) Food Stamps per capita



(c) Share of women receiving welfare



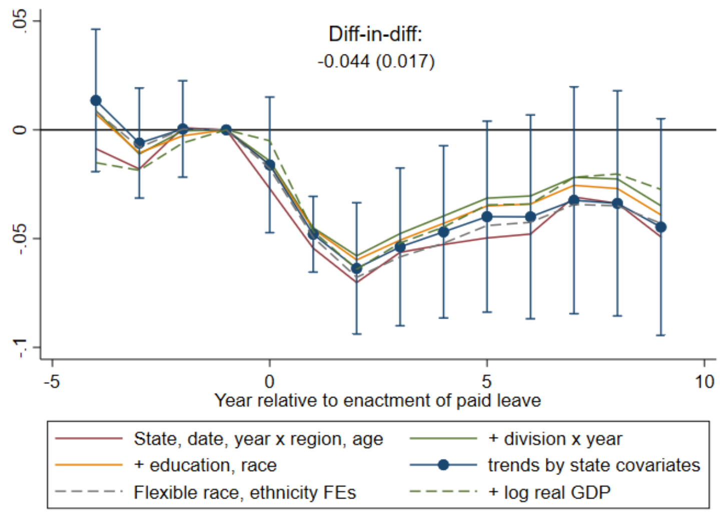
(d) Share of women with any govt income



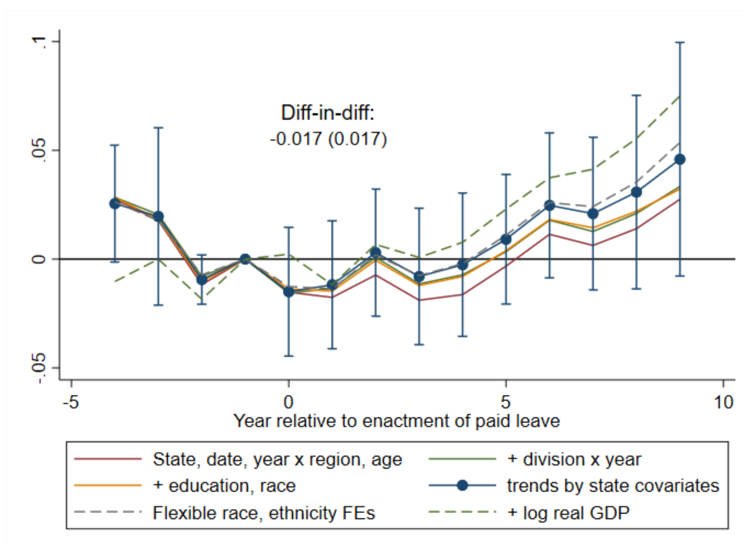
Notes: Panels show estimates of τ from equation (1.6) using measures of transfer income per capita constructed using data from the BEA Regional Income Division and population counts from the National Cancer Institute or data from the March CPS, 1968-1984, accessed via IPUMS (Ruggles et al., 2017).

Figure 1.5: Effects of paid leave on hourly wages

(a) Women age 18-45



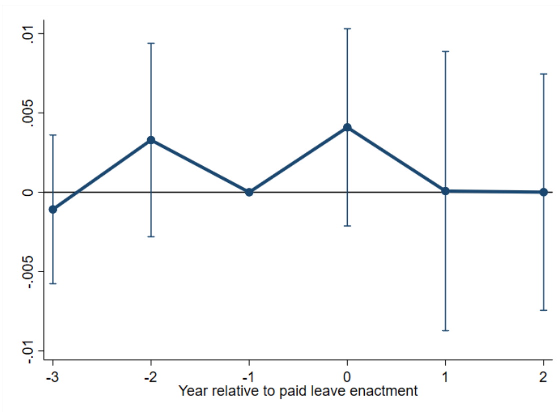
(b) Men age 18-45



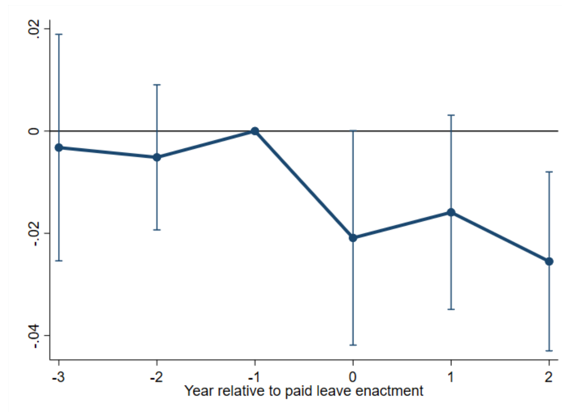
Notes: Graph shows event-study estimates from equation (1.6) using samples of women and men age 18-45 from the 1973-1987 May CPS and 1979-1987 Merged Outgoing Rotation Group files. Sample excludes self-employed and farm workers, as well as wages greater than $100 or less than $1 in 1979 dollars. Weighted regressions use CPS earnings weights where available, and standard CPS sampling weights from 1973-1978. Basic controls include fixed effects for month and year of the survey, state, and a quadratic in age interacted with indicators of nonwhite race and Hispanic ethnicity. Education controls include a linear term in years of schooling plus indicators for completing high school and college. Standard errors are clustered at the state-group level.

Figure 1.6: Effects of paid leave on family income

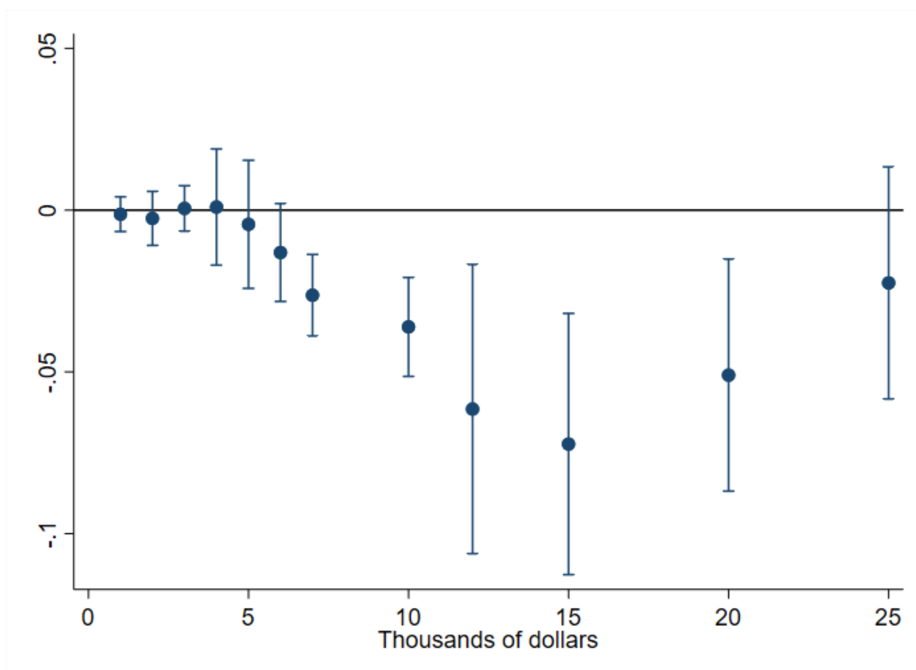
(a) Share with income ≥ $1,000



(b) Share with income ≥ $7,500



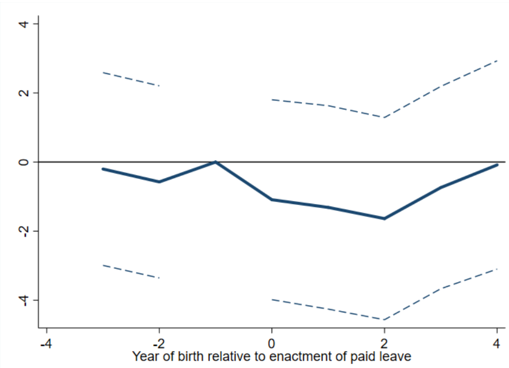
(c) Effect on the distribution of family income



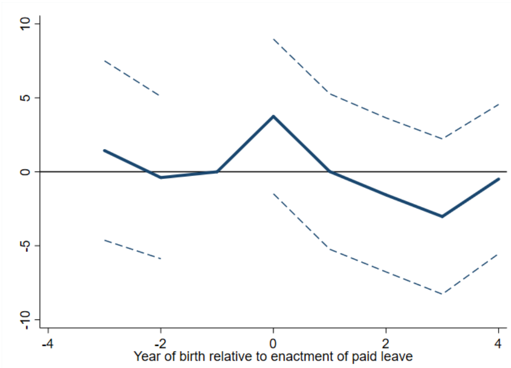
Notes: Figures 1.6a and 1.6b show event-study estimates of the effect of the enactment of paid leave on the share of women age 18-45 in families with income greater than $1,000 and $7,500, respectively. Figure 1.6c shows difference-in-difference estimates of the same effect at various thresholds of family income. Sample includes women age 18-45 from the 1974-1981 May CPS who are the head or wife of the household head. Weighted regressions use CPS earnings weights. Standard errors are clustered at the state-group level.

Figure 1.7: Effect of STDI maternity benefits on fertility

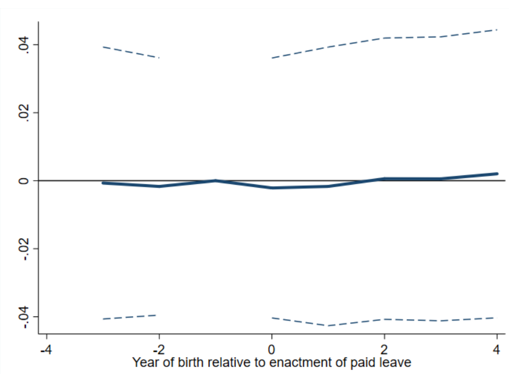
(a) Fertility rates



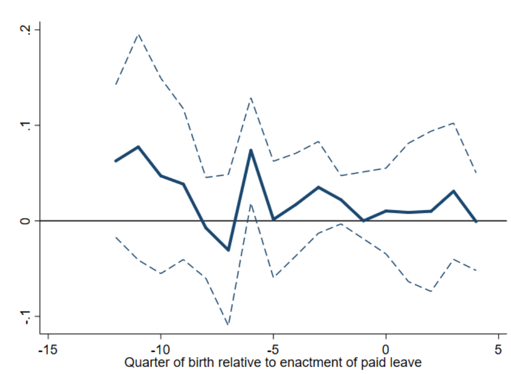
(b) Average birthweight



(c) Share born to nonwhite mothers

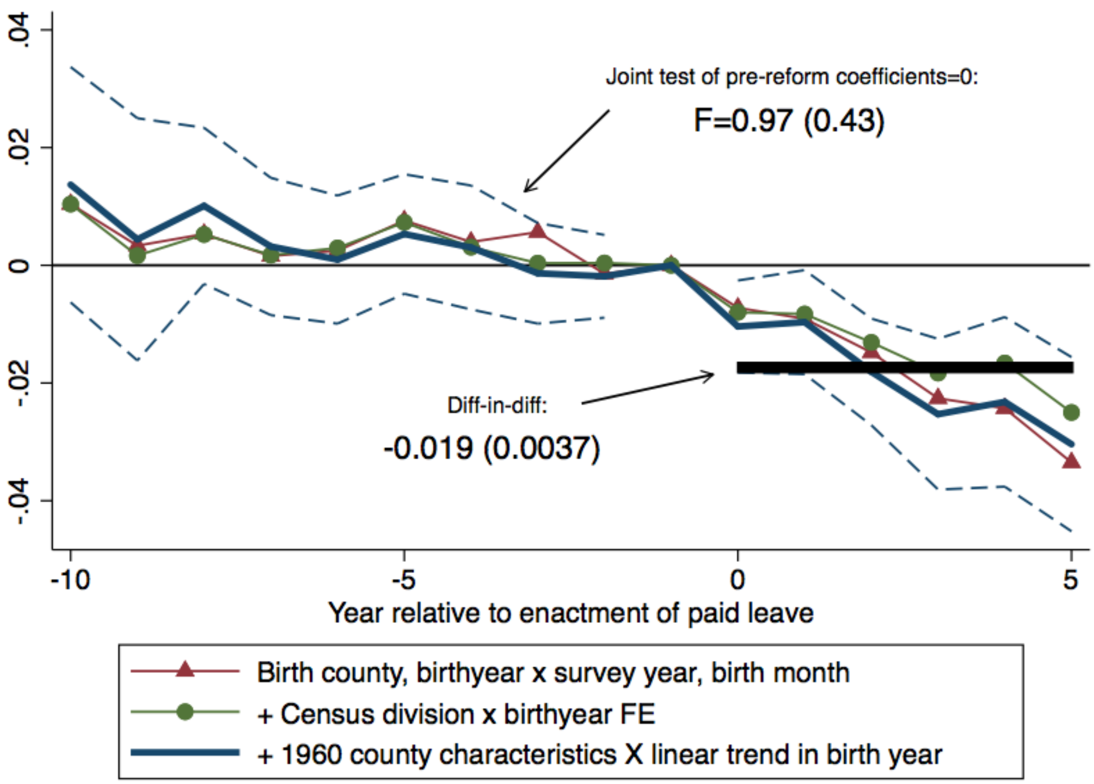


(d) Mothers' education



Notes: Estimates in Figures 1.7a, 1.7b, and 1.7c use birth record data from the Natality Detail File, 1974-1984, accessed via ICPSR, and population counts by age, sex, and race from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. In Figure 1.7d, data on mother's education comes from 1980 long-form decennial Census accessed via IPUMS (Ruggles et al., 2017). Standard errors are adjusted for heteroskedasticity. In Figure 1.7d, standard errors are also adjusted for intracluster correlation within states and individual mothers.

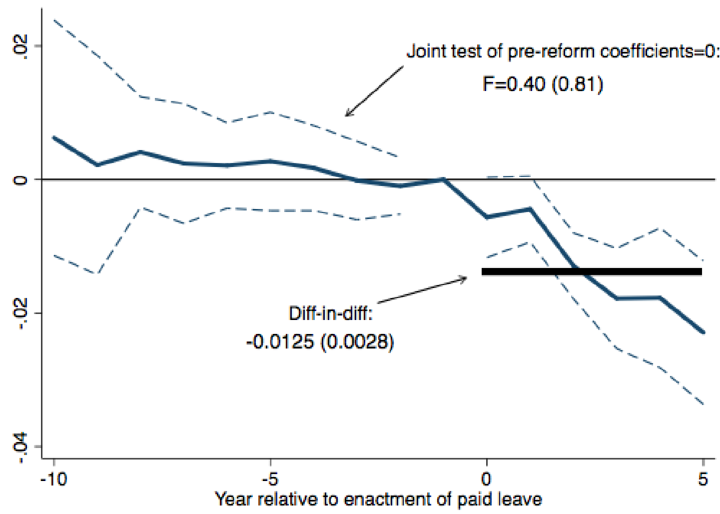
Figure 1.8: ITT effect of paid leave enactment on index of educational outcomes



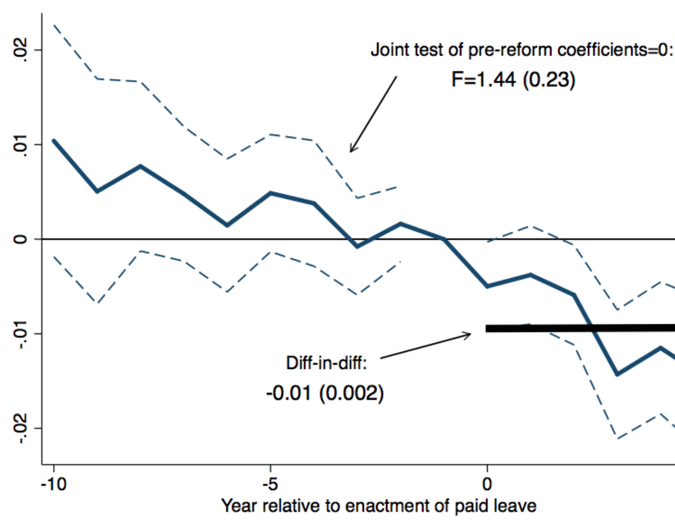
Notes: Coefficients displayed are estimated intent-to-treat effects of exposure to paid maternity benefits on children in the long run. Standard errors clustered at state level. Sample includes individuals in the restricted 2000 long-form decennial Census and 2001-2016 American Community Survey, using cohorts born in the United States between 1954-1985, and individuals age 25 or older when surveyed. Estimated using equation (1.6) with $STDI_{s,1970}$ calculated using industry-level STDI coverage shares from Autor et al. (2013) and 1970 decennial Census microdata (Ruggles et al., 2017).

Figure 1.9: The long-run effects of STDI maternity benefits on children's education

(a) College attendance



(b) College completion



Notes: Coefficients displayed are estimated intent-to-treat effects of exposure to paid maternity benefits on children in the long run. Standard errors clustered at state level. Sample includes individuals in the restricted 2000 long-form decennial Census and 2001-2016 American Community Survey, using cohorts born in the United States between 1954-1985, and individuals age 25 or older when surveyed. Estimated using equation (1.6) with $STDI_{s,1970}$ calculated using industry-level STDI coverage shares from Autor et al. (2013) and 1970 decennial Census microdata (Ruggles et al., 2017).

CHAPTER II

Paid Maternity Leave and the Gender Wage Gap

2.1 Introduction

The erosion of the gap between women's and men's outcomes in the labor market is one of the most important economic stories of last century. While only 20 percent of American women with a child under age 1 reported formal employment in 1970, the share with a job had tripled by the end of the century. At the same time, wages among female workers rose substantially, although they remained well short of the wages of their male counterparts.

Economists have proposed a number of explanations for the narrowing of the gender wage gap – as well as the incomplete nature of this convergence in women's and men's wages – since the mid-20th century. These explanations include the erosion of differences in human capital (e.g., Black et al., 2008; Blau and Kahn, 2017), differential preferences for shorter and more flexible hours (e.g., Goldin, 2014), changes in the selection-induced bias in observed wage distributions (e.g., Blau and Beller, 1988; Mulligan and Rubinstein, 2008), non-cognitive factors such as a reluctance to negotiate or behave competitively (e.g., Niederle and Vesterlund, 2007; Card, Cardoso and Kline, 2016), and the persistence of discrimination (e.g., Goldin and Rouse, 2000).

Intertwined with each of these explanations is the role of public policy and other labor-market institutions. Employers have increasingly adopted practices, both willingly and at the behest of policymakers, that are designed to help workers balance the competing demands of family and the workplace. These policies are generally thought to encourage labor-force participation, reduce career interruptions, and benefit the long-run earnings of working parents – and especially women, given their continued status as the primary providers of home production. However, family-friendly policies may also lead to shifts in demand for women's labor services that feed back in the form of lower wages, fewer job offers, and other deterioration in women's labor-market prospects. While well-established economic theory predicts that certain policies have the potential to create unintended consequences, these effects have proven difficult to document empirically (Olivetti and Petrongolo, 2017; Rossin-Slater, 2017).

This paper explores the effect of an expansion of paid maternity leave benefits on the gender wage gap. In the 1970s, a number of U.S. states, and eventually the federal government, enacted laws that barred employment discrimination against women on the basis of pregnancy. One of the consequences of these laws was that short-term disability insurance, or STDI, was required to cover childbirth as a disability, essentially creating America's first paid maternity leave policy. Because the enactment date of these laws differed across states, and because the share of working women with access to STDI varied from state to state, they created variation in both the timing and intensity of the treatment. This variation provides an opportunity for an empirical exploration of the ambiguous theoretical implications of a substantial expansion of paid leave; while such a policy could in theory increase women's labor-force attachment and earnings potential, it also has the potential to discourage the hiring and promotion of women among firms concerned about the cost of absent employees.

I first explore the effect of paid maternity leave on several measures of equality in the labor market. The gender wage gap was largely flat during the 1970s but narrowed during the 1980s. Using an event-study design, I show that the expansion of paid leave increased the measured wage gap between men and women by about 4 percent. A simple decomposition exercise suggests that the expansion of maternity leave can account for -31% of the convergence between men's and women's wages between 1975 and 1985 – that is, without the enactment of paid leave benefits, women's hourly wages would have risen an additional 2 percentage points relative to men.

A change in the measured gap between the hourly wages of men and women can be the result of changes in the way firms compensate workers for their labor services, or because of changes in the composition of the labor force. Which factors were most important for explaining the effects of STDI-funded maternity leave? I explore this question by analyzing changes in women's labor-market outcomes among several margins. I find no robust evidence of shifts in the amount of time women spent working, either on the extensive or intensive margins. However, consistent with previous literature suggesting that family-friendly policies such as maternity leave may accelerate sorting into particular occupations (Blau and Kahn, 2013; Thomas, 2018), I find that women are less likely to hold a professional or management position in the wake of the reform, while men increase entry into these occupations at a rate of nearly 1-to-1.

These results contribute to several bodies of literature. First, they add additional context to the voluminous literature on the drivers of the gender wage gap in the United States (Blau and Kahn, 1997, 2006; Goldin, 2014; Blau and Kahn, 2017). While previous research has closely examined the role of fundamental factors such as changes in the wage structure, differences in human capital, and labor-market discrimination, we have less empirical evidence on the role played by public policy.

These findings also contribute directly to the literature that seeks to understand the consequences of parental leave policies.[1] Most research on the labor-market effects of these policies has studied reforms in Europe or other OECD countries with generous social safety nets, while less evidence is available from the United States, where policy variation is scarce. The policy examined in this paper consisted of a modest but broadly applied maternity policy that, from the perspective of working women, looked

[1]Two excellent reviews of the literature on parental leave can be found in Olivetti and Petrongolo (2017) and Rossin-Slater (2017). I summarize some key findings from this literature in section 2.2.2.

very much like most of the national paid-leave policies that are under consideration in Congress today.

A robust body of literature has shown convincingly that mothers value the opportunity to take leave from work, and that such leaves have the potential to lead to benefits such as reduced stress, healthier behaviors, and more generous mother-infant bonding periods (Rossin, 2011; Carneiro, Løken and Salvanes, 2015). The results of this paper suggest that these policies also come with measurable costs in terms of women's earnings and career prospects that should be considered in any analysis of future policy. These costs have been acknowledged by analysts and policymakers in many cases (Mathur et al., 2017). However, a careful consideration of the factors that drive the results in this paper suggest that the designs most commonly proposed for paid leave – such as using gender-neutral allotments of time off or financing that does not rely on employer mandates – may not be enough to avoid costs to women's labor-market prospects entirely.

2.2 Background: Disability insurance, maternity leave, and the gender wage gap

The latter half of the 20th century was marked by an unprecedented shift in the role of women in the labor force. As late as the 1950s, the widespread practice of "marriage bars" limited women's opportunities to work (Goldin, 1988), and as late as 1971, at least four states retained laws restricting women's employment just before and after childbirth (Koontz, 1971). The rise of female labor force participation fundamentally altered this dynamic, as advocates of women's rights began to promote policies that would allow new mothers to take time away from work without sacrificing their jobs or paychecks.

This paper studies the effect of parental leave policies on the labor market by exploiting an expansion of access to paid maternity benefits that occurred, in most U.S. states, in the 1970s. The history of this expansion, which ultimately required group

STDI policies to cover childbirth as a "disability," is detailed in Section 1.2. As I will describe more fully in Section 2.3, this reform created two sources of variation – differential timing of adoption of the anti-discrimination laws that required STDI to cover childbirth, as well as differential STDI coverage rates that gave the policy more "bite" in some states than in others – that I will use to study the causal effect of the policy.

This expansion of paid leave provides a rare opportunity to evaluate the effect of a large-scale expansion of paid leave on women's labor-market outcomes in the U.S. context. A key feature of the expansion of STDI maternity leave is that these benefits were relatively modest, providing between one-half and two-thirds of usual wages for 6-10 weeks, depending on the policy in question. In addition, they were bestowed on women who often had no access to maternity benefits, just as millions of working women in the United States have no maternity leave benefits even today. Despite these relatively low benefit levels, take-up of the program was robust, with roughly 50 percent of eligible mothers taking up benefits in universal states in the early years, and an increase of about 5-6 weeks in time spent away from work for women who received the maternity benefits (Timpe, 2019).

The enactment of STDI benefits is also of interest because it occurred on the doorstep of the United States' most dramatic period of gender wage convergence. I will discuss this context more fully in the next section.

2.2.1 The convergence of female and male labor-market outcomes

A healthy literature has tracked the dramatic progress in women's labor-market outcomes, one of the most consequential developments of the labor market in the late 20th century (Goldin, 1994). The 1970s and 1980s were a time of particularly active progress toward gender equality.

The general trends in the labor-market outcomes of women age 18-45 during this period are characterized by Figure 2.1a. In the 1970s, labor-force participation accelerated slightly, even compared to the steady growth of the previous decade, and continued to climb in the 1980s. By 1990, female labor-force participation had grown by about half over its 1968 level. The solid blue line suggests that women were also gaining status; between 1968 and 1990, the share of working women in professional or management occupations, as defined using the Census Bureau's occupation coding scheme, also grew by nearly 60 percent.

Even as women were joining men in the labor force in ever-increasing numbers, progress in women's hourly wages came more slowly. The dashed green line in Figure 2.1a shows the trend in the gender wage ratio, which is interpreted as the average wage among women divided by the average wage among men. In 1968, an hour of female labor earned only about two-thirds of the wage earned by an hour of male labor; the gender wage ratio held steady at this level throughout the 1970s, despite women's gains in participation and occupational status. However, the 1980s marked a period of significant convergence, and the wage ratio among full-time, full-year workers rose to about 0.76 by 1990.[2]

Figure 2.1b provides suggestive evidence that the gender wage ratio responded to the enactment of STDI maternity benefits. I construct the wage ratio separately among two groups of states – those that enacted anti-discrimination laws in late 1976 or early 1977, and those that adopted STDI benefits two years later. I then regress the wage ratio in each group of states on a quadratic trend in calendar year and plot the residuals.[3] The resulting graph shows when the wage ratio was rising unusually

[2] The CPS May and Outgoing Rotation Group (ORG) files provide a direct measure of hourly wage observations beginning in 1973. To facilitate an estimate of the gender wage ratio beginning in the late 1960s, I instead use the CPS Annual Social and Economic Supplement (ASEC) in Figure 2.1a. Hourly wages in the ASEC are measured by dividing last year's wage earnings by the product of weeks worked last year and hours worked last week. This measure is relatively noisy and biased toward workers who are less connected to the labor market. Nevertheless, in years where both series are available, they produce similar trends in hourly wages and the gender wage gap. All subsequent analyses involving hourly wages in this paper will use the more direct measure in the May/ORG files.

[3] Figure 2.1b constructs the wage ratio using full-time female workers age 18-45 and full-time male workers age 18-64, in the spirit of abstracting from issues of convex returns to hours worked (Goldin, 2014) and comparing the wages of women of child-bearing age to the "counterfactual" wage distribution. However, the figure is qualitatively similar regardless of the choice of age range of men, restriction placed on work hours, and a linear time trend instead of quadratic.

fast, and when it was lagging the long-run trend. For each group of states, robust growth in the gender wage ratio dipped substantially for 2-3 years after enactment of STDI maternity benefits before returning to roughly the previous trend. This pattern provides informal evidence that the enactment of maternity benefits may have played a part in the delayed convergence of the gender wage ratio in the 1970s.[4]

What explains these trends in the relative wages of working American women? An active literature has explored several potentially complementary drivers of wage convergence. One factor that has largely worked in women's favor is the evolution of the U.S. wage structure. As demand for manufacturing workers receded and the rise of computers raised the return to "soft" skills thought to be more prominent in women, the wage structure shifted in women's favor (Berman, Bound and Griliches, 1994; Beaudry and Lewis, 2014; Weinberger, 2014).

Other factors that have historically been viewed as important determinants of labor-market wages, such as education and experience (Ben-Porath, 1967; Mincer, 1974; Mincer and Polachek, 1974; Becker, 1975), had more ambiguous effects on the wage gap. Women have closed and even reversed the gap in educational attainment; however, they still disproportionately choose majors or degrees that are rewarded with lower wage premiums (Black et al., 2008; Bronson, 2015). In addition, women have long accumulated lower levels of labor-market experience than otherwise similar men; this gap has narrowed over time but persists (Blau and Kahn, 2017), and much of it appears to be related to child-bearing and the outsized role women continue to play in fulfilling household responsibilities (Fuchs, 1989; Sigle-Rushton and Waldfogel, 2007; Bertrand, Goldin and Katz, 2010; Chung et al., 2017; Kleven, Landais and Søgaard, 2018; Pan et al., 2018; Kleven et al., 2019). In addition to more frequent interruptions in labor-market activity, which can hinder employment opportunities through

[4]It is worth noting that other explanations for the late convergence in the gender wage ratio have been proposed; Goldin (1994) suggests that the growing labor-force participation of working women took some time to translate into growth in labor-market experience, which in turn helped close the wage gap. This paper does not test the theory of Goldin (1994), but rather offers another, potentially complementary explanation for the patterns evident in the evolution of the gender wage gap.

skill deterioration and greater job-search costs, a body of research has documented the concentration of women in occupations that allow for more flexibility in exchange for lower wages (Goldin, 2014; Cortés and Pan, 2018). Together, these factors suggest that, even as women invest more in their human capital, the convergence of the gender wage gap may be slowed by a labor market that rewards – at least in pecuniary terms – commitment to a steady supply of labor both over the course of a workday and a career.

An additional factor is the role of selection into the workforce. Ideally, analysts would want to consider the distribution of wage offers and its difference across otherwise comparable men and women. Instead, they generally observe only the hourly wage received among individuals who choose to participate in the labor market, a limitation that is particularly crucial for women, whose relationship with the labor market has changed in fundamental ways in recent decades. The empirical literature has produced widely varying results on this topic, with influential papers arguing that selection has played a key role in the changing wage gap (Mulligan and Rubinstein, 2008) while others suggest it is less important in research that relies on different data and methodologies (Blau and Kahn, 2006).

Another potential source of wage differentials is the existence of discrimination in the labor market. The importance of this factor is notoriously difficult to measure; while it is well-established that the wage gap cannot be explained by covariates in most data sets, and that women are less likely to reach higher-paying, more prestigious rungs on the job ladder (Goldin, 2014; Card, Cardoso and Kline, 2016), discriminatory factors are difficult to separate from other explanations such as preferences for flexibility, as noted above, or the reluctance to compete with men that has emerged as a new topic of the literature on the gender wage gap (Niederle and Vesterlund, 2007, 2008, 2011). Nevertheless, several researchers have obtained evidence from compelling laboratory or natural experiments that employers are more reluctant to hire or promote women than otherwise comparable men (Neumark, Bank and Van Nort, 1996; Goldin and Rouse, 2000; Correll, Benard and Paik, 2007; Moss-Racusin et al., 2012; Reuben, Sapienza and Zingales, 2014).

2.2.2 The role of family-friendly policies

Each of the factors outlined above can interact with the institutional features of a labor market to alter the measured gap in hourly wages between women and men. These effects may be particularly pronounced for "family-friendly" workplace policies. Such policies may alter women's career expectations and, in turn, raise the perceived return to investments in human capital (e.g., Goldin and Katz, 2002; Goldin, Katz and Kuziemko, 2006; Bailey, Hershbein and Miller, 2012); encourage employee-absence-averse employers to engage in statistical discrimination; alter the selection decision by making employment more flexible; and, to the extent that they increase labor force participation among workers who value flexibility, shift the occupational distribution of female workers.

How do these factors apply to maternity leave policies? Much of the literature on parental leave focuses on the implications for working women who must trade off time at home with a newborn child against the desire to remain in the workforce or preserve her current job match (Klerman et al., 1997). In the absence of maternity leave policies, some women will choose to leave the workforce rather than settle for the length of leave their employers are willing to allow, despite the loss of income, the greater deterioration of general and firm-specific human capital, and the job-search costs they will absorb upon a future return to the labor force. In such a setting, a guaranteed allotment of leave will entice women to preserve a job match. This may preserve their place on the job ladder, increase their future potential earnings, and even lead to greater human capital investment, all forces that should serve to reduce gender disparities.[5].

In fact, a large body of literature suggest there is scope for such positive effects. Re-

[5]This literature generally abstracts from choices about the timing or quantity of children to bear. However, it is worth noting a small literature that finds that the availability of parental leave does play a role in these decisions (Hoem, 1993; Averett and Whittington, 2001; Lalive and Zweimüller, 2009; Farré and Gonzalez, 2019).

search on the take-up of maternity leave policies suggests women value such policies highly and take advantage of the opportunity to take a leave of absence with a newborn when it is made available (Waldfogel, 1999; Berger and Waldfogel, 2004; Baker and Milligan, 2008; Washbrook et al., 2011; Rossin-Slater, Ruhm and Waldfogel, 2013; Byker, 2016; Dahl et al., 2016). The structure of the leave allotment also appears to matter (Lalive et al., 2014); for example, unpaid leave appears to disproportionately benefit more-educated mothers while paid leave is more impactful for relatively disadvantaged women (Han and Waldfogel, 2003; Han, Ruhm and Waldfogel, 2009; Byker, 2016). Interestingly, gender-neutral or parental leave allotments have much smaller (although still positive) impacts on the length of time men spend away from work after fathering a child (Han and Waldfogel, 2003; Baum and Ruhm, 2016; Bartel et al., 2018).

The literature has also generated empirical evidence, albeit less robust, that parental leave policies increase job retention among mothers. The existence of such an effect is likely to be a necessary condition for the existence of a positive overall impact of parental leave policies on women's labor-market opportunities. Social scientists have found positive, but sometimes only suggestive, evidence that reforms in the U.S. and Canada increase mothers' attachment to the labor force (Ruhm, 1998; Berger and Waldfogel, 2004; Hofferth and Curtin, 2006; Baker and Milligan, 2008; Rossin-Slater, Ruhm and Waldfogel, 2013; Byker, 2016; Baum and Ruhm, 2016; Bana, Bedard and Rossin-Slater, 2018). However, such findings are not universal; for example, expansions of the already-generous leave allotments in Europe and other OECD nations appear to have no effect on the subsequent labor-market experiences of women (Lalive et al., 2014; Asai, 2015; Dahl et al., 2016).

Parental leave policies' impact on the labor market become less clear when we consider the potential for broader impacts on the labor market, where it can alter decisions not only about labor supply but also labor demand, human capital accumulation, and other factors discussed in section 2.2.1.

A simple theoretical framework would first consider changes in female labor supply and demand for female labor in a static framework (Summers, 1989; Gruber, 1994). As STDI maternity benefits are enacted, employed women gain a non-wage benefit that would be expected to induce more women to enter the labor market, raising employment rates but also lowering hourly wages as female workers become less scarce. At the same time, firms may respond to the cost of the benefits and the longer, more frequent absences by reducing demand for female labor. Together, these factors lead to ambiguous effects on female employment but a clear prediction that female wages should fall.

We might expect the effects on women's wages to be even more pronounced given more nuanced considerations of the effects of selection, discrimination, and women's preferences for flexibility in the workplace. If the women attracted to the labor force are those who most value flexibility, we would expect an even larger effect on the gender wage gap. First, the new entrants are more likely to experience large gaps in experience that lead them to earn lower wages on average. In addition, given the entrants' preference for flexibility, the mechanism outlined by Goldin (2014) may become more pronounced as women sort into lower-paying, less time-intensive occupations. Furthermore, these effects need not impact only mothers or women who plan to bear children; from the employer's perspective, imperfect information about workers' future child-bearing habits will encourage statistical discrimination that discourages hiring or promotion of women of child-bearing age (Thomas, 2018). Indeed, Blau and Kahn (2013) outline just such an explanation for the greater occupational segregation and lower occupational status of women in developed countries with more generous social safety nets than that in the United States.

While the theoretical underpinnings of these impacts on women's labor-market opportunities is well understood, researchers have struggled to find ways to test them empirically (Olivetti and Petrongolo, 2017; Rossin-Slater, 2017). The United States' national policy, the Family and Medical Leave Act (FMLA), had a relatively small impact on access to job-protected leave. Estimates of its effect on employment and wages produced null estimates (Baum, 2003), although Thomas (2018) reports a decrease in women's likelihood of being promoted. Using the more recent expansion of paid leave in California, two papers find evidence of negative employment effects for women (Das and Polachek, 2015; Sarin, 2017). Asai (2019) focuses on a Japanese reform that decreased the direct cost of paid leave to firms and found that firms responded by hiring more women and paying a higher starting wage.

The relative lack of variation in U.S. parental leave policies has also inspired several papers that use innovative approaches and alternative settings to measure the costs of leave-taking. Gallen (2018) examines a 2002 policy change in Denmark and finds evidence of small negative effects on some co-workers' earnings and suggestive evidence of a decrease in firm survival. In another study using Danish data, Brenøe et al. (2018) exploit the insight that, from the point of view of a firm with many employees, the timing of a worker's pregnancy is plausibly exogenous. They analyze firms' response to maternity leave and conclude the costs are small.

While these studies offer compelling evidence that costs are small, it is unclear how easily we can extrapolate such findings to the United States, where the labor market may not absorb an increase in leave-taking as readily as those in Europe and other OECD countries (Blau and Kahn, 2013). Given the scarcity of variation in parental leave policies in the United States, this question has been difficult to test, but the broad nature of the expansion of STDI benefits provides such an opportunity.

2.2.3 What did maternity leave cost?

How large would we expect the costs of a maternity leave policy to be? In the simple model of mandated benefits outlined Summers (1989), the calculation is simple: the competitive labor market ensures that the cost is passed directly to the worker. In the special case where the value to workers is the same as the cost to the employer, this pass-through is 1-to-1.

One simple way to evaluate the likely magnitude of such an effect is consider how a profit-maximizing hiring manager, facing the prospect of hiring a woman of child-bearing age with uncertain fertility preferences, would calculate the expected cost of STDI benefits paid out for childbirth. Using a sample from the CPS ASEC (Ruggles et al., 2017), I estimate that the fertility rate among 18- to 45-year-old female wage and salary earners in the early and mid-1970s ranged from 45 to 65 per thousand women in the early and mid-1970s.[6] Depending on the generosity of an STDI policy – which varied from as low as 6 weeks of benefits at a 50 percent wage replacement rate to 10 or more weeks at a two-thirds replacement rate at the high end – this calculation suggests the expected cost would have been roughly one-half of 1 percent of a female employee's salary. However, the above calculation likely represents a lower bound on the cost of such a policy. First, it assumes there is no substitute for female workers in the labor market; in fact, the effect on wages may be more pronounced if firms are able to hire men in place of women. In addition, although fertility rates fell after 1970, the fertility rate among *working* women did not fall as dramatically, and was actually rising during the late 1970s, suggesting that hiring managers may have had reason to believe leave-taking among their workers would become more common. Finally, in many cases, the nature of the anti-discrimination laws would have ushered in not only cash benefits for childbirth, but also required employers to continue health insurance benefits for mothers of newborns – a factor that would have added another 0.25 percent of annual wages to the expected cost of hiring a female worker (Baum, 2003; Gruber, 1994).

Other potential costs require the consideration of wage-setting models that go beyond the static, perfectly competitive setting. For example, employers argued vociferously in public debates at the time that they would be forced to do so absorb other costs, such as the hassle costs of hiring temporary replacement workers, as well as the training costs and lower productivity among those replacements (Rousmaniere, 1977; U.S. House of Representatives, 1977; U.S. Senate, 1979). In addition, although ma-

[6] Wage and salary earners are defined as those with at least $100 in wage and salary earnings in the previous year. I drop individuals with business income below 0 or above $100. The fertility rate is the share of this sample with a child age 0 in the household.

ternity leave policies are designed to promote job continuity among mothers, they may also increase turnover in the labor market overall, either by inducing selection into the workforce by less-attached women or by altering fertility or occupational choices.

While difficult to quantify, such costs could be substantial. Using Swiss administrative data, Blatter, Muehlemann and Schenker (2012) estimate that average hiring costs are equivalent to 10-17 weeks of wage earnings. The bulk of this cost is attributed to the lower productivity of new workers and the training required to bring them up to speed; these "adaption costs" alone account for 71 percent of the hiring.

2.3 Data and Empirical Strategy

To investigate the labor-market consequences of the expansion of paid maternity benefits through STDI, I construct a sample of men and women age 18-64 from the May Current Population Survey (CPS), 1973-1978, and the Outgoing Rotation Group of the CPS from 1979-1993. The chief strength of these data is that they are the earliest available data to directly measure hourly wages (or weekly earnings and usual weekly hours worked) in large, regularly collected, nationally representative samples from the United States.[7] In addition, I observe employment status, industry and occupation, and basic demographic characteristics of the respondents.

My primary focus is on women of "child-bearing age," or 18-45 for the purposes of this paper, as well as men of the same age. For women in this age group, the advent of STDI maternity benefits may have represented an important benefit that altered

[7]Between 1973 and 1976, the public version of the CPS does not specify the state of residence for some individuals, assigning them instead to a larger "state group." These state groups differ from the state groups used in the March CPS available through IPUMS; specifically, several of the 27 states identified individually in the CPS data are instead grouped in the IPUMS data. It is unclear whether this discrepancy is due to different coding schemes or a coding error, so I recode all variables to the more conservative IPUMS coding scheme. In cases where states within a "state group" adopted the STDI maternity benefit reform at different times, I include the respondents in the sample but assign them to the omitted event-time category for all years. As a result, these individuals do not contribute directly to my estimates of interest but do increase their precision by contributing to my estimates of fixed effects and other covariates. My results do not change significantly if I use the alternative CPS state coding scheme or drop states with conflicting treatment dates.

labor-supply decisions and human capital accumulation. In addition, to the extent that firms responded to the expected or actual costs of maternity leave, it is reasonable to suspect that these groups would have been the most likely to experience a change in the demand for their labor services.

To evaluate the effect of maternity leave on the labor market, I make use of two sources of variation: The varying time of adoption of anti-discrimination legislation within a state, and the varying "bite" of this legislation across states with differing pre-existing levels of STDI coverage. Because data on STDI coverage at the state level is rare, particularly during the time frame of my study, I construct my best estimate by combining industry-level STDI coverage rates with data on the distribution of working women across industries in 1970. The construction of this measure, which I denote $STDI_{s,1970}$, is described in Section 1.4.1.

My main approach is to use this estimate of the "bite" of the maternity leave policy, $STDI_{s,1970}$, to augment an event-study design that allows me to flexibly estimate the impact of exposure to maternity benefits:

$$y_{ist} = STDI_{s,1970} \sum_{k \neq -1} \tau_k 1\left\{k = t - T_s^*\right\} + \boldsymbol{\delta_s} + \boldsymbol{\theta_{r(s)t}} + \boldsymbol{X'_{ist}\beta} + \epsilon_{ist} \tag{2.1}$$

where $\boldsymbol{\delta_s}$ is a set of state fixed effects, $\boldsymbol{\theta_{r(s)t}}$ is a fixed effect at the Census-region-by-year (or division-by-year) level, $\boldsymbol{X'_{ist}}$ is a suite of covariates, and y_{ist} is an outcome of interest such as the log of the hourly wage or an indicator for being employed in a professional or management occupation. The coefficients of interest, $\boldsymbol{\tau_k}$, flexibly estimate the intent-to-treat effect of expansion of maternity benefits on outcome y_{ist}. Because $STDI_{s,1970}$ scales my event-time dummies in a way that proxies for the share of women who received maternity benefits as a result of the anti-discrimination laws, these estimates can be interpreted as the effect of making maternity leave benefits universal.

In addition to specification 2.1, because I am particularly interested in women's *relative* wages, I also use a specification that pools men and women and adds an interaction term that estimates the effect of the policy on women's wages, net of any impacts on hourly wages for all workers, as follows:

$$
\begin{aligned}
y_{ist} = {} & STDI_{s,1970} \sum_{k \neq -1} \psi_k Female_{ist} 1\left\{k = t - T_s^*\right\} \\
& + STDI_{s,1970} \sum_{k \neq -1} \gamma_k 1\left\{k = t - T_s^*\right\} + \boldsymbol{\delta_s} + \boldsymbol{\theta_{r(s)t}} + \boldsymbol{X'_{ist}} \boldsymbol{\beta} + \epsilon_{ist}
\end{aligned} \tag{2.2}
$$

where in this case my parameter of interest is ψ_k, which can be interpreted as the intent-to-treat effect on women's wages relative to those of men.

In each of these specifications, the key identifying assumption is that no other determinants of y_{ist} are correlated with the expansion of STDI maternity benefits. For example, a recession-led deterioration in women's labor-market outcomes could confound my estimates, but only if it affected workers in a given state at the same time maternity benefits were enacted, and in such a way that the effect was correlated with the pre-existing share of female workers who received STDI. To account for such factors, I adjust for state-level output using data from the Bureau of Economic Analysis' Regional Economic Accounts data, as well as other potential determinants of wages and other labor-market outcomes. In particular, because STDI coverage appears to be correlated across states within a region, I include $\boldsymbol{\theta_{r(s)t}}$ to adjust flexibly for any trends that may be common to states in a particular region of the country.

2.4 Results

This section reviews the results of my estimates of the effect of the expansion of STDI maternity benefits on women's labor-market opportunities, and particularly the gender wage gap. I first discuss my estimates of the effect and quantify their magnitude in terms of the broad trends in women's relative wages. I then move on to the channels through which maternity benefits access led to these changes in women's wages, and their implications for the experiences of American working women.

2.4.1 The effect of STDI benefits on the gender wage ratio

The main result of the paper is pictured in Figure 2.2. The estimate of each element of $\boldsymbol{\tau_k}$ is plotted on the vertical axis against the event-time on the horizontal axis. The flexible event study design provides a built-in test for a pre-trend; the coefficients are statistically indistinguishable from 0. However, the women's relative wages break sharply after the enactment of STDI benefits, leveling out at about 4-5 percent below the previous trend. As shown in the figure, these estimates are remarkably consistent across a range of specifications.

What does this decrease in relative wages mean for the gender wage gap? Figure 2.3 provides a visual representation of the estimated effect on the gender wage ratio – interpreted as the ratio of the hourly wage of the average working woman and the hourly wage of the average working man – in a subset of states for which I can observe the treatment date in public CPS data.[8] The observed gender wage ratio, shown as the dashed line, is broadly consistent with Figure 2.1a. Women's hourly wages were roughly two-thirds of average male wages in the 1970s, but converged quickly in the 1980s, reaching nearly 80 percent of male wages by the end of the decade. The counterfactual is represented by the solid line, which represents the gender wage ratio that would have been observed in the absence of STDI maternity benefits. The two lines can be distinguished statistically at the 5 percent (10 percent) level in years with triangle (circle) markers, but are indistinguishable in other years. The two lines

[8] While my estimates of the change in female and male wages use all individuals that meet my sample criteria to improve precision, for this exercise I drop individuals from several states where, due to the grouping of states in CPS data, I cannot accurately measure event time. These states include Wisconsin, Michigan, Maine, New Hampshire, Vermont, Rhode Island, Iowa, Missouri, Minnesota, North Dakota, South Dakota, Nebraska, Kansas, Hawaii, Alaska, Oregon, Washington, Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Georgia, Virginia, Delaware, Maryland, West Virginia, and South Carolina. In addition, I drop New Jersey and Connecticut because they adopted STDI maternity benefits before my sample period.

begin to diverge in 1978 and grow further apart over the time period.

The year-by-year estimates of the gender wage gap can be further evaluated in Table 2.1. Column 1 shows the estimated gender wage ratio, while column 2 shows the wage ratio in the counterfactual scenario in which states did not enact laws require group STDI plans to cover childbirth. Column 3 shows the difference between the two estimates, as well as the p-value from a test of the null hypothesis that the counterfactual wage ratio is no larger than the observed ratio. My estimates are noisy in the early years of the sample period, a feature that can be attributed in large part to the smaller sample sizes before 1979. However, beginning in 1978, we can reject the null that the STDI maternity benefit reform did not restrict the convergence of the gender wage ratio.

How large was this effect? A convenient reference point is 1975, when my estimates of the actual and counterfactual gender wage ratio are nearly identical, at 0.69. Over the next 10 years, women gained just over 6 percentage points on men. However, the ratio would have closed another 1.9 percentage points under the counterfactual scenario, suggesting that STDI maternity benefits can explain -31 percent of the convergence in women's relative wages over this period. Even though women's relative wages continued to grow substantially through the end of the 1980s, the persistence of the estimated effect of STDI maternity benefits meant that the gap between the counterfactual and measured wage ratio also continued to grow; this exercise suggests -25 percent of the converge from 1975-1990 is explained by STDI benefits.

2.4.2 The effects of STDI maternity benefits on labor supply and women's occupational distribution

The stark effects of the expansion of STDI maternity benefits on women's relative wages raise additional questions about the consequences of parental leave policies for women's labor-market opportunities. To what extent did these changes stem from changes in selection into the workforce or shifts in the occupations in which women were working?

To the extent that the reform initiated changes in women's willingness to work or the demand for female labor, we might expect to see effects on the time women spent on the job. Table 2.2 shows no such effects. Here I have pooled event time in equation 2.1 into three-year groups to provide summary results in table form. Column 1 provides an estimate of changes on the intensive margin by using the log of usual hours worked; the estimate of a 2 percent increase in the first years after the reform cannot be distinguished from 0.

In some cases, small or nonexistent shifts in average hours worked may mask changes at other points in the distribution. To explore this possibility, I also regress indicators on the likelihood of working more than 0, 20, or 40 hours on equation 2.1. Column 2 effectively provides an estimate of the extensive-margin response of labor supply; here again I see no statistically significant change in the likelihood of working positive hours. Similarly, columns 3 and 4 show no evidence of an effect at either the 20-hour or 40-hour margin.

Finally, one potential response to the reform is that women may shift across industries. The implied direction of such an effect is not clear, since firms with a history of offering STDI benefits may be particularly reluctant to hire women of child-bearing age, even as women seek out positions at those firms. Column 5 of Table 2.2 tests this question by regressing an indicator for employment in a high-STDI industry on equation 2.1; again I see no significant effect.

While the results above suggest that STDI benefits did not have an impact on the overall quantity of female labor supplied or the industries in which women worked, the estimates in Table 2.3 suggest that shifts in occupation did indeed play a role. Column 1 provides a baseline estimate of the effect on the gender wage gap, suggesting that women's relative wages fell by about 4 percent in the first six years after the enactment of STDI maternity benefits. Column 2 shows that little changes if we add a proxy for the macroeconomic environment.

However, the addition of occupation fixed effects significantly cuts the estimated effect of the reform – by nearly one-third in the first few years after the reform and nearly half in years three through five. This result suggests that a significant portion of the decrease in female relative wages came as a result of shifts into lower-paying occupations. Such an effect is consistent with previous literature that has observed that working women tend to choose jobs that provide flexibility as a tradeoff for high pay Goldin (2014), and that family-friendly policies may reinforce this effect by encouraging such sorting behavior (Blau and Kahn, 2013).

Such occupational segregation effects have typically been difficult to measure empirically. However, one natural explanation is that, in the wake of the enactment of STDI maternity benefits, women may have been less likely to be promoted to high-paying, prestigious occupations (Thomas, 2018). Figure 2.4 shows estimates of a regression of an indicator for being employed in a professional or management occupation – as determined by the harmonized OCC1990 variable provided by IPUMS (Ruggles et al., 2017) – on equation 2.1.[9] In general, these results are less robust than the effects on women's relative wages, with noisy estimates in the period prior to the reform that may suggest a positive pre-trend. However, Figure 2.4a shows a pronounced negative shift in the share of women with professional or management occupations, a trend break that is underestimated if in fact there is a positive pre-trend. Relative to the year before the reform, the share of women in professional and management occupations fell by about 1 percent – or by about 9 percent relative to the control mean of 13 percent. The persistence of the effect even 10 years later is reminiscent of the lasting change in female relative wages observed in Figure 2.2. At the same time, men age 18-45 appear to experience a similarly sized *increase* in the share with a professional or management occupation, although this effect is less persistent. These mirror-image effects are consistent with sorting of men and women into different occupations, even in the absence of aggregate effects on employment or hours worked, and in a way that

[9] To construct this outcome variable, I merge the IPUMS crosswalk for the OCC1990 variable to the CPS codes of both 1970 and 1980 vintages. I then use OCC1990 codes for "managerial and professional specialty occupations" (3 through 200) as a proxy for a "high-prestige" position. Non-employed individuals are coded as not in a professional or management occupation, although the results are similar if I focus on the sample of employed individuals.

hindered the convergence of the gender wage gap throughout the 1980s.

2.5 Conclusion

How would the expansion of parental leave benefits impact the U.S. labor market? The answer to this question has proven elusive, largely because of a lack of policy variation in the United States (Olivetti and Petrongolo, 2017; Rossin-Slater, 2017).

This paper seeks to shed light on this question by evaluating a large-scale expansion of maternity leave benefits through STDI. This policy provided modest pecuniary benefits – and, at least in some cases, some protection against layoff – for working mothers. Importantly, the payments provided by these benefits are similar in magnitude to those suggested by prominent proposals for national paid leave today (Lee and Ernst, 2019; DeLauro and Gillibrand, 2019).

My findings suggest that although parental leave policies are often intended to expand labor-market opportunities for women, they also have the potential to create unintended consequences that exacerbate inequality gender. My estimates suggest that the enactment of STDI maternity benefits created a persistent wedge between men's and women's wages, and that the historic gender wage convergence in women's that occurred between 1975 and 1990 would have been 25 percent larger in the absence of the policy.

My results also suggest that a large part of this decrease in women's relative wages can be attributed to women sorting into lower-paying, less prestigious occupations. My estimated effects on relative wages fall by one-third to nearly half when relying on variation within occupation, and the enactment of maternity benefits appears to have led to a 9 percent drop in the share of women working in a professional or management occupation.

The most recent several decades have seen most developed countries dramatically expand access to maternity and paternity leave for new parents; in the United States, which has lagged far behind its peers in this arena, the push for a national paid leave policy enjoys unusually broad support. The experience of STDI maternity benefits suggests policymakers should use caution when designing such policies. The evidence suggests that parental leave policies do in fact carry costs along with the benefits, that these costs exist even when the pecuniary benefits are not remitted by the firm, and that these costs may not be borne only by those who use the policy but rather by all women in the workforce.

Table 2.1: Estimated effects of STDI maternity benefits on the evolution of the gender wage ratio

Year	(1) Observed wage ratio	(2) Counterfactual wage ratio	(3) Difference	(4) Share explained
1975	0.692	0.693	-0.001	
	(0.009)	(0.013)	[0.46]	
1976	0.703	0.707	-0.004	-38%
	(0.011)	(0.015)	[0.36]	
1977	0.705	0.704	0.001	9%
	(0.010)	(0.014)	[0.56]	
1978	0.693	0.701	-0.008	-485%
	(0.008)	(0.011)	[0.05]	
1979	0.704	0.720	-0.015	-124%
	(0.007)	(0.016)	[0.06]	
1980	0.716	0.732	-0.015	-63%
	(0.008)	(0.014)	[0.04]	
1981	0.723	0.738	-0.016	-51%
	(0.009)	(0.014)	[0.03]	
1982	0.735	0.750	-0.015	-36%
	(0.009)	(0.016)	[0.05]	
1983	0.747	0.763	-0.017	-31%
	(0.010)	(0.016)	[0.04]	
1984	0.749	0.765	-0.016	-28%
	(0.007)	(0.017)	[0.08]	
1985	0.755	0.774	-0.020	-31%
	(0.010)	(0.020)	[0.07]	
1986	0.760	0.783	-0.023	-33%
	(0.008)	(0.020)	[0.09]	
1987	0.773	0.797	-0.024	-30%
	(0.008)	(0.021)	[0.08]	
1988	0.784	0.811	-0.027	-29%
	(0.009)	(0.022)	[0.07]	
1989	0.795	0.822	-0.027	-26%
	(0.009)	(0.023)	[0.09]	
1990	0.804	0.832	-0.028	-25%
	(0.010)	(0.024)	[0.09]	

Notes: Data includes men and women age 18-45 from the 1973-1978 May CPS and 1979-1993 CPS Outgoing Rotation Group files from NBER. Gender wage ratio is calculated as the exponential of the difference between average log wages for women and men in each year. Counterfactual is calculated as the average predicted log wage with estimates of ψ_k from equation 2.2 set to 0. Individuals from states that adopted STDI maternity benefits before 1975 have been dropped from the wage ratio calculation, as have individuals from states where the date of STDI maternity benefit adoption cannot be observed due to CPS state groupings. Standard errors (in parentheses) are calculated using a clustered bootstrap with 1,000 replications. P-value (in brackets) is obtained from a one-sided test of the null hypothesis that the counterfactual gender wage ratio is not higher than the observed gender wage ratio. The share explained is the ratio of the difference between the observed and counterfactual wage ratios and the difference between the ratios observed in 1975 and a given year.

Table 2.2: No effects on hours worked or industry

	(1)	(2)	(3)	(4)	(5)
	Log hours worked	Works>0 hours/week	Works>20 hours/week	Works>40 hours/week	High-STDI industry
Event years -4 to -2	0.0154	0.00895	0.0115*	-0.00535*	0.00749
	(0.0136)	(0.00614)	(0.00649)	(0.00273)	(0.0116)
Event years 0 to 2	0.0212	0.00148	0.00587	-0.00389	-0.00903
	(0.0153)	(0.00704)	(0.00870)	(0.00335)	(0.0102)
Event years 3 to 5	0.0238	0.00899	0.0156	-0.00524	0.00687
	(0.0201)	(0.0110)	(0.0143)	(0.00455)	(0.0107)
Observations	600,494	1,063,681	1,063,681	1,063,681	539,171
R-squared	0.027	0.020	0.019	0.009	0.008
Control mean	32.7	0.572	0.487	0.0523	0.311

Notes: Data includes men and women age 18-45 from the 1973-1978 May CPS and 1979-1993 CPS Outgoing Rotation Group files from NBER. Estimates shown come from equation 2.1 with event time pooled into three-year groups. All models include controls for race, ethnicity, age, state per-capita GDP, state fixed effects, and Census-division-by-year fixed effects. Standard errors (in parentheses) are clustered at the state group level.

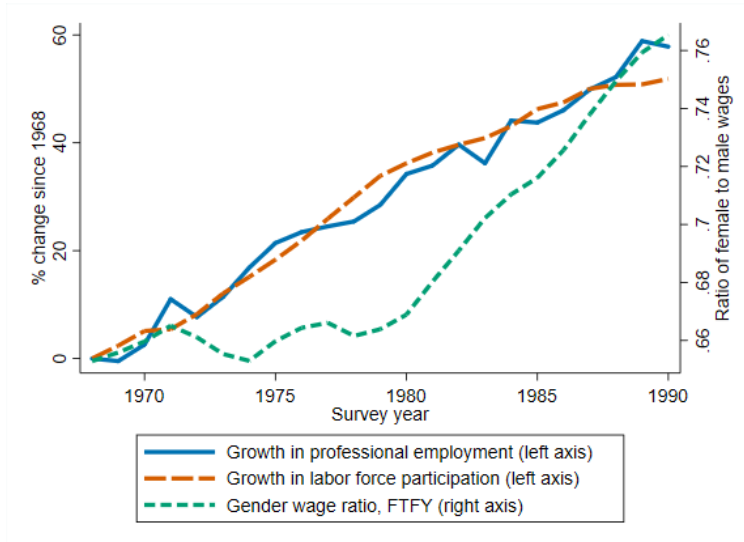
Table 2.3: Effect of STDI on women's relative wages: The role of occupation

	(1) Log wages	(2) Log wages	(4) Log wages	(5) Log wages
Event years -4 to -2	-0.0206	-0.0188	-0.0151	-0.0140
	(0.0139)	(0.0134)	(0.0114)	(0.0117)
Event years 0 to 2	-0.0407***	-0.0392***	-0.0281***	-0.0270***
	(0.00988)	(0.0102)	(0.00554)	(0.00628)
Event years 3 to 5	-0.0382**	-0.0366**	-0.0201**	-0.0220**
	(0.0164)	(0.0170)	(0.00863)	(0.00873)
Age, race, ethnicity	X	X	X	X
Per-capita output		X	X	X
Occupation			X	X
Education				X
Observations	1,258,257	1,258,257	1,258,257	1,258,257
R-squared	0.324	0.324	0.492	0.510

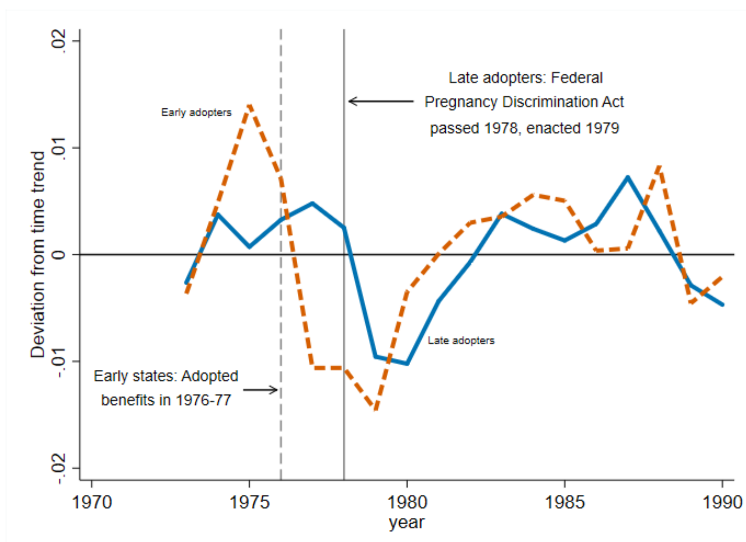
Notes: Data includes men and women age 18-45 from the 1973-1978 May CPS and 1979-1987 CPS Outgoing Rotation Group files from NBER. Estimates shown are ψ_k from specification 2.2 with event-time grouped into three-year intervals. Standard errors are clustered at the state group level.

Figure 2.1: Women's employment, hourly wages, and the advent of STDI maternity benefits

(a) Evolution of female employment and the gender wage ratio

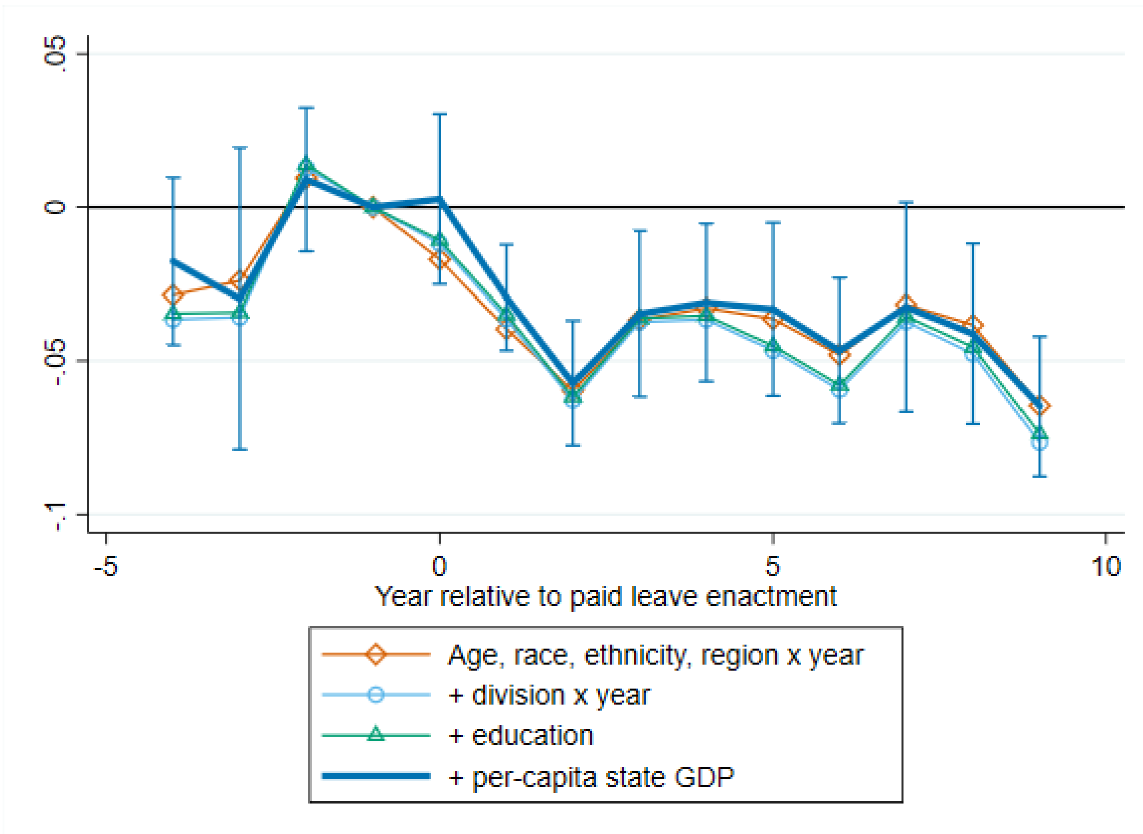


(b) Wage convergence slows after maternity benefits adopted



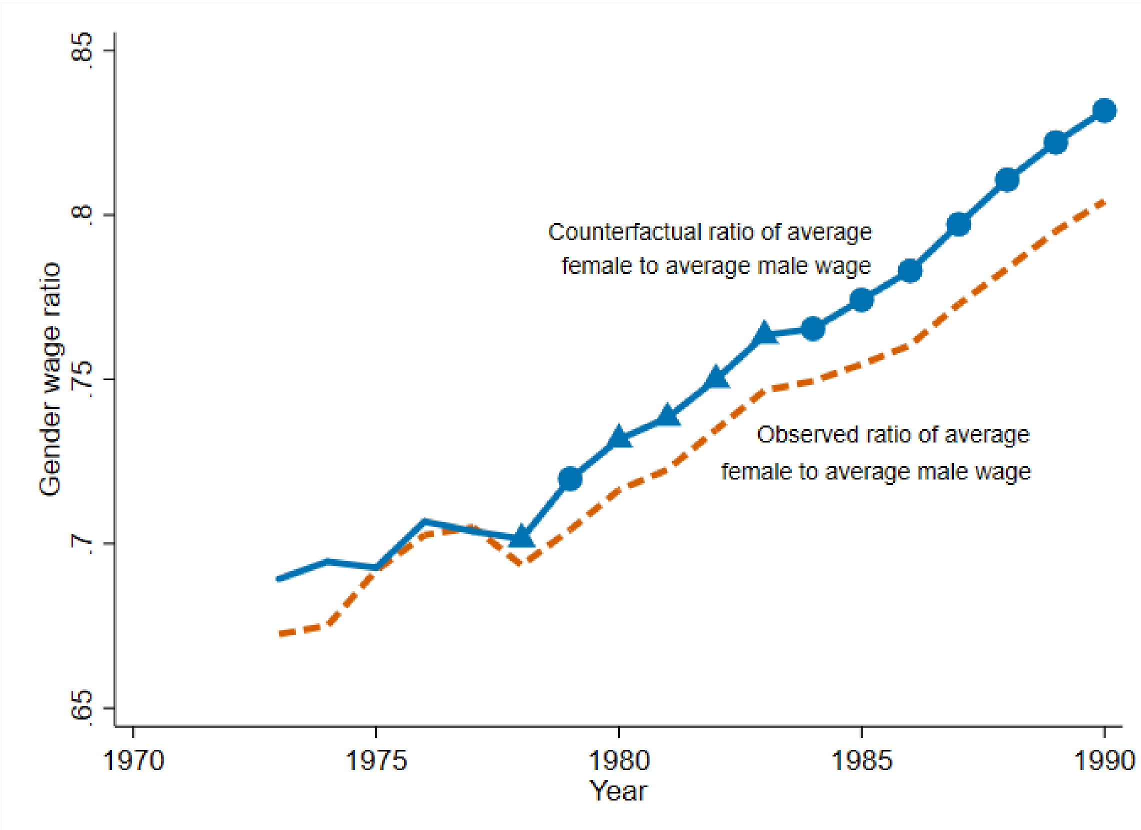
Notes: Data in Figure 2.1a comes from the 1968-1999 CPS Annual Social and Economic Supplement (Figure 2.1a) and 1973-1993 CPS May and MORG files (Figure 2.1b). Figure 2.1b plots the deviation from trend of the gender wage ratio for full-time workers, separately for two groups of states: One that adopted STDI maternity benefits in late 1976 and early 1977, and one that adopted benefits after passage of the Pregnancy Discrimination Act of 1978. The gender wage ratio is calculated as the exponential of the difference in the average log hourly wage for women age 18-45 and the average log hourly wage for men age 18-64 in each year. Deviation from trend is calculated as the residual from a regression of the gender wage gap for full-time employees on a quadratic time trend. Labor-force participation and professional employment (defined using Census codes for professional and management occupations) are calculated among women age 18-45.

Figure 2.2: Paid leave leads to decrease in women's relative wages



Notes: Graph shows event-study estimates from equation 2.1 using samples of women or men age 18-45 from the 1973-1987 May CPS and MORG files. Standard errors are clustered at the state-group level.

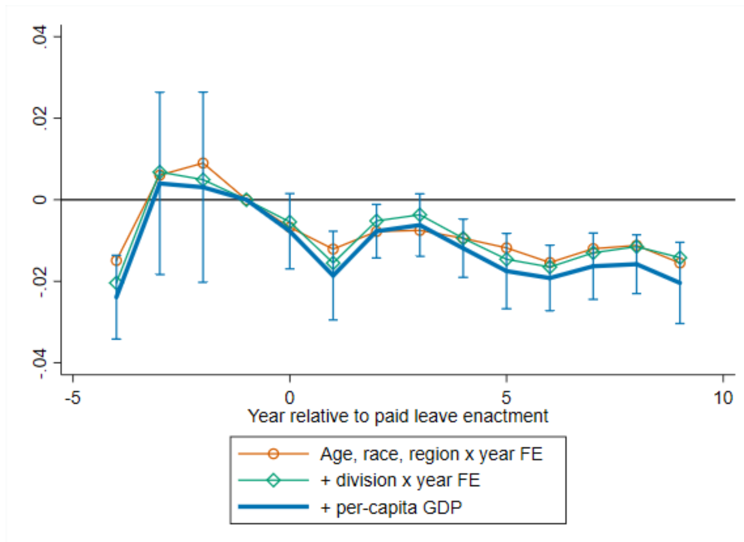
Figure 2.3: Impact of enactment of STDI maternity benefits on wage convergence



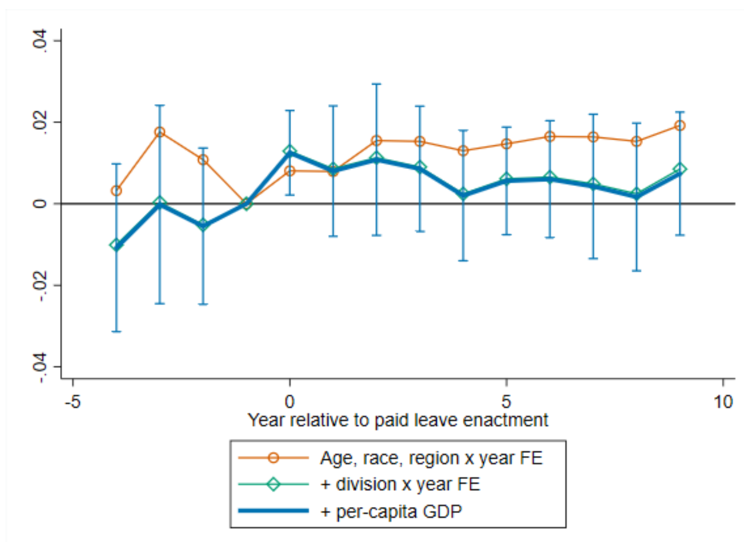
Notes: Sample is drawn from 1973-1987 May CPS and MORG files from NBER. Gender wage ratio is calculated as the exponential of the difference between average log wages for women and men in each year. Counterfactual is calculated as the average predicted log wage with estimates of ψ_k from equation 2.2 set to 0. Individuals from states that adopted STDI maternity benefits before 1975 have been dropped from the wage ratio calculation, as have individuals from states where the date of STDI maternity benefit adoption cannot be observed due to CPS state groupings. Estimates of the counterfactual gender wage ratio with blue triangle (circle) markers are statistically distinguishable from the observed gender wage ratio at the 0.05 (0.10) level. Confidence intervals for the test of the differences between the counterfactual and observed gender wage ratios are calculated using a clustered bootstrap with 1,000 replications.

Figure 2.4: Paid leave and sorting into professional and management occupations

(a) Women



(b) Men



Notes: Figures show event-study estimates from equation 2.1 using samples of women or men age 18-45 from the 1973-1987 May CPS and CPS MORG files. Per-capita GDP is constructed by dividing state output figures from the Bureau of Economic Analysis by population counts from the Surveillance, Epidemiology, and End Results program at the CDC. Professional and management occupations are defined as codes 3 through 200 in the IPUMS OCC1990 variable (Ruggles et al., 2017). Non-employed individuals are coded as not in a professional or management occupation, regardless of their reported occupation. Standard errors are clustered at the state-group level.

CHAPTER III

Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency

> Convincing evidence on the longer-term impacts of scaled-up pre-k programs on academic outcomes and school progress is sparse, precluding broad conclusions. The evidence that does exist often shows that pre-k-induced improvements in learning are detectable during elementary school, but studies also reveal null or negative longer-term impacts for some programs.

– Brookings Pre-Kindergarten Task Force of Interdisciplinary Scientists, (Phillips et al., 2017)

3.1 Introduction

In 1965, the U.S. began a new experiment in the provision of public preschool for disadvantaged children. The motivation was simple: "the creation of and assistance to preschool, day care, or nursery centers for 3- to 5-year-olds . . . will provide an opportunity for a head start by canceling out deficiencies associated with poverty that are instrumental in school failure" (United States Senate Committee on Labor and Public Welfare, 1964). The program that ensued is the now-famous "Head Start," a "prep school for poor kids" which aimed to help millions of children escape poverty (Levitan, 1969).

More than fifty years later, Head Start is one of the most popular of the War on Poverty's programs, serving around 900,000 children annually at a cost of $9.6 billion in 2017 to the federal government. Unlike expensive, small-scale "model" programs such as Perry Preschool and Abecedarian, Head Start's architects prioritized widespread access, calculating that a massive preschool expansion would maximize its poverty-fighting (and political) benefits. Skepticism about the quality of this large-scale preschool program coupled with difficulties in evaluation have generated controversy about its short-term benefits for decades (Duncan and Magnuson, 2013; Currie, 2001; Westinghouse Learning Corporation, 1969). Convincing evidence regarding Head Start's long-term effects has remained even more elusive, thanks to the lack of program randomization in its early years, small sample sizes of longitudinal surveys, and the difficulty of measuring adults' access to Head Start decades ago. Consequently, the best estimates of Head Start's long-term effects are limited by lingering concerns about endogeneity (sibling comparison designs, Currie and Thomas (1995); Garces, Thomas and Currie (2002); Deming (2009) and imprecision (measurement error in funding and access, Ludwig and Miller (2007); Carneiro and Ginja (2014). Whether Head Start achieved its goal of increasing life opportunities for children remains an open question. This paper uses large-scale data to estimate Head Start's long-term effects on human capital and economic self-sufficiency. By linking the restricted long-form 2000 Census and 2001-2013 American Community Surveys (ACS) to the exact date and place of birth from the Social Security Administration's (SSA) Numident file, we observe outcomes for one-quarter of U.S. adults as well as a high quality measure of their access to and eligibility for Head Start as children. The resulting sample is four orders of magnitude larger than longitudinal surveys, and information on place of birth and exact date of birth ameliorates (potentially non-classical) measurement error in childhood access to Head Start.

Our research design exploits the county-level roll-out of Head Start programs from

1965 to 1980 at the Office of Economic Opportunity (OEO) (Levine, 1970; Bailey and Goodman-Bacon, 2015; Bailey and Duquette, 2014; Bailey and Danziger, 2013; Bailey, 2012). This approach exploits the well-documented "great administrative confusion" at the OEO (Levine, 1970), mitigating problems of measurement error in archival funding data (Barr and Gibbs, 2017) and concerns about the endogeneity of Head Start funding levels. An additional strength of our design is that it leverages Head Start's age-eligibility guidelines, comparing cohorts who were age-eligible when it launched (ages 5 and younger) to cohorts born in the same county that were age-ineligible (children 6 and older). The substantial variation in adult outcomes means that even our large dataset does not permit the estimation of a regression discontinuity or regression kink design, but our approach is based on similar assumptions. Much like a regression kink design (Card et al., 2015), our key identifying assumption is that Head Start's causal effect is the only reason for a change in the relationship between a child's age at the program's launch and her outcomes as an adult. Because even our large-scale data are not large enough to estimate a regression kink formally, we present both event-study estimates as well as trend-break tests based on spline parameterizations.

The results suggest that Head Start increased the human capital and economic self-sufficiency of disadvantaged children. An index of adult human capital rose by 10 percent of a standard deviation among Head Start participants relative to children born in the same county who were age 6 when the program began. Participating children achieved 0.29 more years of education, were 2.1 percent more likely to complete high school, 8.7 percent more likely to enroll in college, and 19 percent more likely to complete college. In addition, Head Start increased economic self-sufficiency in adulthood by almost 4 percent of a standard deviation—gains driven largely by a 12-percent reduction in adult poverty and a 29-percent reduction in public assistance receipt. We find no evidence of reductions in incarceration. Heterogeneity in Head Start's effects suggests that they reflect, in part, practices outside of pre-school curriculum. In particular, health screenings and referrals as well as more nutritious meals appear to be important mechanisms for the program's effects on disadvantaged children. In addition, the effects of Head Start appear to be complement greater family and public resources arising from a stronger economy. Overall, Head Start appears to have achieved the goals of its early architects, both increasing children's economic opportunities and reducing poverty.

A final analysis quantifies the private, internal rates of return to dollars spent on Head Start in the 1960s and 1970s. Rather than using changes in wage income directly, we use the National Longitudinal Survey of Youth 1979 (NLSY79) to predict changes in earnings for the relevant cohorts net of any ability differences (Neal and Johnson, 1996; Deming, 2009). Using potential earnings accounts for Head Start-induced negative selection in men's employment (driven by reductions in disability) and positive selection in women's employment (due to the income effect dominating for the least skilled women). This exercise suggests a private internal rate of return to Head Start of 7.7 percent, which ranges from around 4 percent for women to 11 percent for men. Using only savings on public assistance expenditures as a conservative method to calculate the program's social returns, we find that the internal rate of return of putting one child through Head Start is 2.4 percent. In short, this paper's estimates suggest substantial long-run returns to America's first scaled-up, public preschool program. While the results do not imply that all of today's large-scale preschool programs work, they suggest that some less-than-model preschool programs may have lasting effects—a key finding for current policy deliberations (Phillips et al., 2017).

3.2 The Launch of Head Start in the 1960s and Expected Effects

In the 1960s, the idea that preschool could improve children's cognitive development was revolutionary. Challenging the conventional notion that IQ was immutable and fixed at birth, Joseph McVicker Hunt's (1961) book, Intelligence and Experiences, persuasively argued that children's intelligence could be significantly improved by altering their experiences. Benjamin Bloom further emphasized that the first four years of children's lives was a "critical period," noting that "intelligence appears to develop as much from conception to age 4 as it does from age 4 to 18" (Bloom, 1964). This idea suggested an innovative strategy for poverty prevention. Because poor children started school with significantly less educational background, comprehensive preschool could give them a "Head Start," improving their success in school and addressing a root cause of poverty.

3.2.1 A Brief History of Head Start's Launch

Funded by the OEO, Head Start began as an 8-week summer program in 1965. After a successful first summer, President Lyndon Johnson announced that Head Start would become a full-year program for children ages 3 to 5. The director of the OEO wrote 35,000 letters to public health directors, school superintendents, mayors and social services commissioners to encourage applications. The OEO also made a special effort to generate applications in America's 300 poorest counties (Ludwig and Miller, 2007).

Head Start's political popularity led to an even faster launch than other War on Poverty programs. Figure 3.1 shows the program's quick expansion. By 1966, Head Start had begun in more than 500 counties where over half of the nation's children under age 6 resided. By 1970, federal expenditures on the program reached \$326 million, or \$2.1 billion in 2018 dollars (OEO, 1970). This early expansion ensured that by 1970, Head Start existed in roughly half of U.S. counties, putting preschool programs within a short drive for 83 percent of children under age six (Appendix Table B.1).

The exact timing of Head Start's launch depended on many idiosyncratic factors. The OEO's "wild sort of grant-making operation" has been well documented in oral histories (Gillette 1996: 193) as well as in more recent, quantitative analyses (Bailey and Duquette, 2014). In the case of Head Start, other factors were key as well: how excited were local institutions or politicians about the program? Was their there adequate and available space to launch? Could the program be integrated within the public school system or would it remain separate? The final result of the grant-making process and local constraints was that Head Start began earlier in areas that were significantly more populous and urban (Appendix Table B.2)—areas where more children could be served. In addition, urban areas were funded earlier. After accounting for population and urban differences, the roll-out of Head Start was not strongly affected by other county characteristics (Appendix Table B.3). Consistent with the historical evidence that this national program was rushed into existence, exactly when Head Start began after 1965 does not appear to be systematically related to pre-existing local characteristics.

3.2.2 Head Start's Mission

Head Start's architects adopted a holistic approach that aimed to develop children's mental and physical abilities by improving health; self-confidence; verbal, conceptual, and relational skills; and raising parental involvement. Levitan (1969) notes that Head Start's 1966-7 budget included early childhood education (daily activities and transport, 70 percent), health services (including immunizations, screenings and medical referrals, 4 percent), and nutrition (14 percent). Parent involvement, social services (e.g., helping families cope with crises), and mental health services accounted for the remaining budget.

The expected effects of the program on adult outcomes could flow directly from the early learning facilitated by the program. But the role of health services and nutrition may be important as well. Head Start's vaccinations and screening (e.g., tuberculosis, diabetes, vision, hearing) and referrals to local physicians may have prevented complications from childhood diseases (North, 1979; Ludwig and Miller, 2007) and helped parents obtain simple, cost-effective technologies to improve learning (e.g., eye glasses and hearing aids or antibiotics to reduce hearing damage from ear infections). Healthy meals and snacks may have also raised children's ability to learn. Early estimates suggest that more than 40 percent of children entering Head Start were receiving less than two-thirds of the recommended allotment of iron, and 10 percent were extremely deprived in terms of their daily calories (Fosburg et al., n.d.). Among children who received blood tests in the 1968 full-year program, 15 percent were found to be anemic (Office of Child Development , DHEW). Reducing these nutritional deficiencies could also translate into significant gains in educational achievement in both the short and longer term (Frisvold, 2015).

The challenges of quickly starting a new national program meant that implementation often deviated from ideals. Not only did Head Start lack curricular standardization, but programs struggled to find high-quality teachers to achieve the suggested pupil-to-teacher ratio of 15:1. As a practical solution many centers relied on para-professionals, most of whom lacked post-secondary education; thirty percent had not finished high school (Hechinger, 1966; Braun and Edwards, 1972). In addition, many components of Head Start phased in slowly. For instance, the OEO wrote that in 1965, "the proportion of children receiving treatment for conditions discovered in Head Start medical and dental examinations ... was probably under 20 percent. It rose to over 65 percent in 1966, and in 1967 we fully expect it to have reached over 90 percent" (OEO, 1967).

Consequently, Head Start in its earliest years was far from a model preschool program. Nevertheless, even the less-than-ideal implementation of Head Start was likely higher quality than the alternatives available to low income children in the 1960s (Currie, 2001). Importantly, similar concerns hold today: Head Start's quality score from the National Institute for Early Education Research places Head Start program quality around the median of the score distribution (Espinosa, 2002) but the program may still be much better than informal child care (Loeb, 2016).

3.3 Literature Regarding the Long-Term Effects of Head Start

Previous evaluations of Head Start provide suggestive evidence of the program's long-term effects on human capital and economic self-sufficiency. One pioneering approach was the use of family fixed effects with longitudinal data. Building on work by Currie and Thomas (1995), Garces, Thomas and Currie (2002) used the Panel Study of Income Dynamics (PSID) to compare children who participated in Head Start to their siblings who did not. They show that Head Start increased high school graduation rates and college enrollment among whites and reduced arrest rates among blacks. Using a similar research design for more recent cohorts in the National Longitudinal Survey of Youth (NLSY), Deming (2009) finds that Head Start participation had large and positive effects on a summary index of adult outcomes (including high school graduation, college attendance, "idleness," crime, teen parenthood, and health status). Well-known critiques caution that sibling comparisons may suffer from sources of endogeneity bias (Griliches, 1979; Bound and Solon, 1999). In addition, small sample sizes in longitudinal surveys may provide unreliable estimates of Head Start's effects (Grosz, Miller and Shenhav, 2017).

More recent work exploits shifts in access to Head Start using three distinct research designs. The path-breaking application of RD in Ludwig and Miller (2007) exploited the OEO's special effort to generate grant proposals from the 300 poorest counties. Comparing the outcomes of children on either side of this threshold, they find evidence that Head Start reduced childhood mortality and increased the receipt of high-school degrees and college enrollment. However, because the 1990 and 2000 Censuses required them to use county of residence in adulthood to proxy for childhood Head Start access, measurement error causes their education results to be sensitive to specification and often statistically insignificant. Carneiro and Ginja (2014) use an RD in state-, year-, and household-based income eligibility cutoffs for more recent Head Start programs. They find that Head Start decreased behavioral problems, the prevalence of some health conditions (including obesity) between the ages of 12 and 17, and crime rates around age 20. They find a positive though statistically insignificant effect on receiving a high-school diploma as well as suggestive evidence that Head Start reduced college enrollment.

In work closely related to this paper, three studies make use of county-year variation in Head Start funding in the 1960s and 1970s to quantify the program's long-term effects. Using a sample of likely eligible children from the NLSY, Thompson (2017) finds that greater funding for Head Start at ages 3 to 6 raised college graduation rates, reduced the incidence of health limitations, and tended to raise adult household income. Focusing on a "high impact" sample, Johnson and Jackson (2017) find that an average level of Head Start and education spending increases the likelihood that children graduated from high school by 8 p ercentage points and gained 0.39 years of schooling. These children also experienced a 7.8 log-point increase in adult wages, a 14.4 log-point increase in adult family income at ages 20 to 50, a 3.6 percentage-point reduction in poverty at ages 20 to 50, and a 3 percentage-point reduction in adult incarceration. Finally, Barr and Gibbs (2017) examine the intergenerational effects of Head Start using the NLSY and two research designs: family fixed effects and variation in program availability across birth counties (also referred to as "roll-out"). To alleviate concerns about the endogeneity of funding levels and measurement error in the National Archives data, their roll-out design uses a binary measure of Head Start access that is equal to one if funding exceeds the 10th percentile of observed funding per four-year-old. They find evidence of large first-generation effects on women (including a gain of a half a year of schooling) and large second-generation effects on their children's high school graduation and completed education.

3.4 Data and Research Design

This study combines the long-form 2000 Census and 2001-2013 ACS with the SSA's Numident file to shed new light on Head Start's long-term effects. The Census/ACS data represent almost one quarter of the U.S. population and are four orders of magnitude larger than previously used longitudinal samples. Another advantage of these combined data is that the Numident contains county of birth (rather than adulthood residence) and exact date of birth, which allows a high-quality proxy for Head Start access and age eligibility in childhood. The data's main disadvantage is that they contain no information on family background. This lack of covariates means that we cannot model many determinants of adult outcomes—which limits precision even in this large dataset—or model treatment effect heterogeneity by childhood characteristics.

Our sample is comprised of children born from 1950 to 1980 in U.S. states where the school-entry age cutoff is known. We additionally limit our sample to individuals who are in their prime earning years (ages 25 to 54). We collapse these data to means by birth year, survey year, county of birth, and school age. We also weight our regressions using the number of observations in each cell (Solon, Haider and Wooldridge, 2015). To minimize disclosure concerns at the Census Bureau, we use only observations with non-allocated and non-missing values for all outcomes.

Our outcomes of interest are summary measures of human capital and economic self-sufficiency, which permit tests of co-movements of related adult outcomes and limit the number of statistical tests (Kling, Liebman and Katz, 2007). A shortcoming of this approach is that, because indices weight each component equally, large changes in one dimension are averaged with potentially opposite-signed or zero effects in other dimensions. We, therefore, also examine the individual index components. The human capital index includes four binary variables indicating achievement of a given level of education or greater: high school or GED, some college, a 4-year college degree, and a professional or doctoral degree; years of schooling, and an indicator for working in a professional occupation. Our index of self-sufficiency includes binary indicators of employment, poverty status, income from public sources, family income, and income from other non-governmental sources; continuous measures of weeks worked, usual hours worked, the log of labor income, log of other income from non-governmental sources, and log ratio of family income to the federal poverty threshold.

3.4.1 Measuring Exposure to Head Start

Combining data on the launch of Head Start programs from Bailey and Goodman-Bacon (2015) with the Census/ACS-Numident permits two refinements to previously used research designs Barr and Gibbs (2017); Johnson and Jackson (2017); Thompson (2017). First, we use only variation in the launch of the Head Start program rather than a continuous measure of Head Start spending. This refinement (1) addresses the potential endogeneity of Head Start funding levels to the program's performance and (2) sidesteps issues of measurement error in the National Archives grant data (Barr and Gibbs, 2017). Second, we examine changes in outcomes for children who were age-eligible for Head Start (ages 3-5 or younger) relative to those who were age-ineligible (ages 6+) when it launched, allowing for the effects to vary by the number of years each cohort was potentially eligible. Age eligibility is based on exact date of birth in the Numident and school-entry age cutoffs, which alleviates measurement error in defining the potential treatment and control groups. Finally, our large dataset allows us to use state-by-birth-year fixed effects to adjust estimates for state economic and policy changes that could have affected children's outcomes independently of Head Start. Our identifying assumption in the analysis that follows is that the causal effect of Head Start is the only reason for a change in the relationship between a child's age at the program's launch and her outcomes as an adult. (See Appendix for more description of our sample.)

3.4.2 Event-Study Regression Model

Our research design uses a flexible event-study framework and roll-out of Head Start to estimate the effect of exposure to Head Start on long-term human capital and economic outcomes,

$$Y_{bct} = \theta_c + \alpha_t + \delta_{s(c)b} + \boldsymbol{Z'_c} b\boldsymbol{\beta} + HeadStart_c \left[\boldsymbol{Age'_{bs(c)}\phi} \right] + \epsilon_{bct} \qquad (3.1)$$

Children's birth years are indexed by $b = 1950 - 1980$, county of birth by c, and Census/ACS year by $t = 2000 - 2013$. Specifications include fixed effects for county of birth, θ_c, year, α_t, and state-by-birth-year, $\delta_{s(c)b}$, which, respectively, capture time-invariant differences across counties, national changes affecting all individuals, and changes in state policies that differentially affect birth cohorts. Although covariates matter little, we follow the literature and include county characteristics, Z_c interacted with a linear trend in year of birth, b (Hoynes, Page and Stevens, 2011; Bailey, 2012; Bailey and Goodman-Bacon, 2015).

$HeadStart_c$ is a binary variable equal to 1 if a child was born in a county that received a Head Start grant before 1980. Age is a set of dummy variables for a child's "school age" at the time of Head Start's launch, $1(T_c^* - b = a)$ where $a = -15$ to 30 (or $T_c^* = 1965$ and $b = 1980$ to $T_c^* = 1980$ and $b = 1950$) and T_c^* is the year Head Start began in county c. We omit school age 6 (age 6 before the school entry cut-off date), because these children would have been unlikely to have attended Head Start rather than public school. Our point estimates of interest, ϕ, describe the evolution of the intent-to-treat (ITT) effects of Head Start on long-term human capital and economic self-sufficiency. Standard errors are corrected for heteroskedasticity and adjusted for an arbitrary within-birth-county covariance structure (Arellano, 1987; Bertrand, Duflo and Mullainathan, 2004). In our tables, we also report p-values corrected for multiple hypothesis testing using the Bonferroni-Holm method in our tables (Holm, 1979; Duflo, Glennerster and Kremer, 2007).

3.4.3 Expected Effects of Exposure to Head Start by Age at Launch

The event-study model provides a flexible approach that imposes few restrictions on the relationship between Head Start and adult outcomes. Although economic theory does not make predictions as to the magnitudes of the event-study coefficients, the program's phased implementation and the greater potential for some children to enroll (due to multiple years of exposure) predict a specific pattern. Figure 3.2 plots this pattern under the following set of assumptions.

First, if we assume Head Start had no effect on children who were over age 5 when the program launched, then the relationship between adult outcomes and Head Start for these children should be zero. This is the equivalent to a test for a pre-trend in our analysis and is illustrated as a flat line for children ages 6 to 12 in Figure 3.2a.

Second, if Head Start has a positive causal effect on adult outcomes, we expect to see a change in those outcomes for children under age 5 when it launched, because these cohorts would have been the first to have been age-eligible and have access. This would not result in an immediate shift in the level of outcomes (akin to a regression discontinuity, RD) but rather a shift in the slope (akin to a regression kink, RK). The reason is that Head Start's capacity grew over time, both because new programs were added but also as individual programs matured. Program quality also increased over time with better hiring and training of teachers, curriculum development, and the implementation of auxiliary services (e.g., health). Studies of other War on Poverty programs such as family planning or community health centers suggest that many of these programs reached maturity around 4 to 5 years after launch (Bailey, 2012; Bailey and Goodman-Bacon, 2015). Figure 3.2a illustrates this as the implementation curve (line with square markers), which rises from zero to 100 percent.

In addition to these gradual changes in program quality and capacity, we also expect larger effects for children who were younger when Head Start launched, simply because they would have been age-eligible for a larger share of their preschool years. For instance, a child 5 years old when Head Start launched could participate for at most one year, whereas a 3-year-old child would be age-eligible for three years. This does not mean that the 3-year-old enrolled for more than one year. However, it is more likely that a child enrolls if s/he had three years to do so. Figure 3.2a illustrates this cumulative potential access to Head Start as a linear relationship (dashed line with circle markers), but differences in the likelihood of enrollment by age could make this relationship more S-shaped as well (because enrollment in the early years was more likely at ages 4 than 5).

The combination of phased implementation and cumulative potential access to Head Start implies a non-linear change in the relationship between age at Head Start launch and adult outcomes (solid, bold line). In Figure 3.2a's stylized example, children ages -1 or younger at Head Start's launch would have been age-eligible for a fully implemented program for each of their three years of eligibility. Assuming that Head Start did not continue to mature and that it did not have any complementarities with other War on Poverty programs, the relationship should level off for children ages -1 or younger at launch, because all cohorts born after this would have had the same potential exposure to Head Start as the -1 cohort.

Note, however, that Figure 3.2b shows how relaxing two assumptions implies a slightly different shape. First, allowing for effects on children ages 6 and older implies that the curve would begin to slope up before age 6. This is possible because 10 percent of children in full-year Head Start were 6 or older (Vinovskis, 2005), and age-ineligible children could still benefit from their younger siblings' participation (Garces, Thomas and Currie, 2002). Because our subsequent analysis standardizes the effects at age 6 to zero, this relationship would appear as the flat part of the line falling below zero. Second, if the Head Start program continued to mature after 5 years or was complemented by other programs (e.g., Goodman-Bacon (2018) notes that Medicaid continued to expand into the 1970s), we would expect to see a slope for cohorts ages -1 and younger when the program launched.

3.4.4 Spline Summary Specification

Our event-study estimates impose none of the restrictions used to outline the expected effects of Head Start in Figure 3.2. However, we expect estimates from this flexible event-study specification to be noisy, in part because so many unobserved factors determine adult outcomes. To improve precision and test formally for trend breaks, we use Figure 3.2's predictions to guide the specification of a three-part spline with knots at ages 6 and -1. We implement this by replacing the $Age'_{bs(c)}$ in equation 3.1 with components of the spline in age, $a = T^*_c - b$:

$$Y_{bct} = \theta_c + \alpha_t + \delta_{s(c)b} + \boldsymbol{Z_c'} b\boldsymbol{\beta} + HeadStart_c \left[\boldsymbol{D'_{cb}\rho_1} + a\boldsymbol{D'_{cb}\rho_2}\right] + \epsilon_{bct} \tag{3.2}$$

where D'_{cb} is a vector of dummy variables, $1(-10 \leq a \leq -1)$, $1(-1 \leq a \leq 6)$, $1(6 \leq a \leq 15)$, $1(a < -10)$, and $1(15 > a)$ and the other variables remain as previously defined. We constrain the estimates of ρ_1 and ρ_2 to ensure that the spline joins at $a = 6$ and -1. While the spline specification is more restrictive than our flexible event-study, it improves precision by imposing restrictions that should be true. The spline allows a parsimonious method to test for a pre-trend (captured in the slope of the segment for $1(6 \leq a \leq 15)$) as well as a formal trend-break test between components $1(6 \leq a \leq 15)$ and $1(-1 \leq a \leq 6)$. This final test captures whether the relationship between adult outcomes and age at Head Start's introductions changes at age 6—the age at which older children tend to stop participating in Head Start and begin first grade.

3.5 Tests of Identifying Assumptions

The research design outlined in the previous section relies on two crucial assumptions: (1) the launch of a Head Start program increased participation in Head Start and (2) the launch of a Head Start program did not coincide with other county-level changes that would affect the outcomes of preschool children. This section provides further evidence on both assumptions.

3.5.1 How Much Did Head Start Increase Preschool Enrollment?

While there is little doubt that introducing a Head Start program increased children's attendance in this program, the magnitude of this relationship net of crowd-out is crucial for interpreting the ITT effects recovered in equations 3.1 and 3.2. Administrative data suggest that the launch of a Head Start program significantly increased children's enrollment. The OEO reported that full-year Head Start served over 600,000 children before 1968, rising from 20,000 children in 1965, to 160,000 in 1966, to around 215,000 in 1967 and 1968 (OEO 1965, 1966, 1967, 1968, 1970). About 257,700 children attended full-year Head Start in 1970. Three-quarters of the children were aged 4 or 5, three-quarters were nonwhite, and 62 percent came from families with less than $4,000 in annual income. Between 1971 and 1978, enrollment and directory information suggest that the average county with a Head Start program served roughly 309 children. These sources imply that the average Head Start program served from about 10 percent of resident age-eligible children in 1971 to 15.8 percent in 1978.

If Head Start substituted for private preschool for some children (Cascio and Schanzenbach, 2013; Kline and Walters, 2016; Bassok, Fitzpatrick and Loeb, n.d.), administrative data may overstate the role of Head Start programs in increasing exposure to preschool. To examine this possibility, we use the 1970 Census, which was the first to ask children younger than age 5 about school enrollment as of February 1—a date during the school year, which should capture enrollment in full-year Head Start. The Census data show that four-year-old children in counties without Head Start programs were 3.4 percentage points less likely to be enrolled in school (16.8 versus 20.2 percentage points, see Figure B.1). Five-year-old children were 17 percentage points less likely to be enrolled in school (48.9 versus 65.9 percentage points). These gaps are 5.9 percentage points among 4 year olds and 21.3 percentage points among 5 year olds when looking only at children of mothers with less than a high school education.

We use a linear probability model to adjust these gaps for state fixed effects (to account for age-invariant, state-level factors that determine the local supply of preschools) and 1960 county characteristics (share of county population in urban areas, in rural areas, under 5 years of age, 65 or older, nonwhite, with 12 or more years of education, with less than 4 years of education, in households with income less than $3,000, in households with incomes greater than $10,000, local government expenditures, income per capita, and whether the county was among the 300 poorest counties). Details are presented in Appendix section 6 and summarized here for parsimony. The regression results show that school enrollment was 14.9 percentage points higher for all five year olds, 15.1 percentage points higher for boys, and 14.5 percentage points higher for girls (Appendix Table B.5). These results are highly robust to the inclusion (or exclusion) of different covariates.

Consistent with crowd-out being minimal, the 14.9 percent increase in enrollment in the Census is only slightly smaller than the 15.8 percent contained in administrative data. This estimate is also comparable to other studies. Garces, Thomas and Currie (2002) estimates the national Head Start participation rate was between 10 percent and 17 percent for the 1964 to 1970 cohorts in the PSID (p. 1002). A slightly higher estimate comes from Ludwig and Miller (2007), who estimate that children's enrollment in Head Start was 17 percentage points higher at the 300-poorest county discontinuity in the 1988 National Educational Longitudinal Survey (NELS).

Based on this evidence, we use our best estimate of 0.149 to transform the ITT effects from our event-study and spline specifications into average treatment-effects-on-the-treated (ATET). We also construct confidence intervals using a parametric bootstrap procedure with 10,000 independent draws from normal distributions with means and standard deviations equal to the point estimates and standard errors from the reduced-form and first-stage estimates (Efron and Tibshirani, 1993).

3.5.2 Did Head Start's Launch Correspond to Other Policy Changes?

Another key assumption of our analysis is that the launch of a Head Start program is the only reason why a cohort's adult outcomes changed for age-eligible children relative to 6+ year olds. The decentralization of U.S. governance and the process of applying for OEO grants makes it unlikely that communities across the nation coordinated in their grant applications. More worrisome is that OEO could have provided multiple grants to a community in the same year, making it difficult to disentangle the effects of Head Start from other OEO programs. To test this hypothesis, we use information on other War on Poverty programs compiled from the National Archives and estimate regressions similar to equation 3.1. In particular, we replace the dependent variable with an indicator for receiving funding in county c in fiscal year t, or $Y_{ct} = \theta_c + \delta_{s(c)t} + X'_{ct}\beta + \sum_k^K \phi_k 1(t - T_c^* = k) + \epsilon_{cbt}$. We also include county and state-year fixed effects, θ_c and $\delta_{s(c)t}$, and county level covariates, X_{ct}, as we do in our primary regression specifications. Our variable of interest is event-time, $t - T_c^* = k$, the year of observation relative to the date Head Start launched (the year before Head Start began, $k = -1$, is omitted).

Figure 3.3 shows the relationship between the launch of Head Start and the launch of other OEO programs. As expected, 100 percent of treated counties in our sample first received a Head Start grant in event-year 0. This is by design. In subsequent years, the share of counties receiving a Head Start grant tapers off to 70 percent after around five years—this reflects the fact that some counties received multi-year grants and also the fact that not all of the early programs continued.

For our estimates to be confounded by changes in other federal funding, these funding changes would need to happen around the time Head Start launched, or event-year 0. However, our analysis of Food Stamps, Community Health Centers, and other child health programs show no such pattern. The one program that shows a small change in funding after Head Start began is the CAP health project. Importantly for our inferences, funding was very small for this program (around \$8 in 2013 dollars per person; by contrast, annual Head Start funding was nearly \$1,500 per 4-year-old during the same period). Furthermore, there is little reason to think this program would affect children who were 5 or younger but not those who were 6. Although we cannot rule out funding changes in programs we do not measure, these patterns provide little evidence that our research design is confounded by other OEO programs.

3.6 Head Start's Effects on Human Capital

Figure 3.4 plots the event-study estimates for all outcomes in the human capital index for the set of compositionally balanced county-birth-cohorts, or individuals ages 15 to as young as -1 when Head Start launched. The solid line plots the event-study estimates. Consistent with the patterns anticipated in Figure 3.2, the human capital index and each of its components exhibit little relationship to adult outcomes for cohorts ages 6 to 15 when Head Start launched (i.e., there is little evidence of a pre-trend among ineligible cohorts). However, the index and many of its subcomponents exhibit a trend-break around age 6, suggesting that access to Head Start improved the human capital of adults. Notably, this relationship is not as sharp as in regression kink designs, because (1) school age-entry cutoffs were not strictly enforced, (2) older children ages 6 and 7 participated in Head Start (although at lower rates), and (3) older siblings of participants may have benefited indirectly from their younger sibling's involvement. Any benefits of Head Start for these other ages would lead the event-study estimates for ages higher than 6 to fall slightly below zero, as the data show. They also weaken the visual evidence and formal tests for a trend-break at exactly age 6.

Table 3.1 summarizes both the event-study and spline estimates at -1. Column 1 presents the mean and standard deviation of the outcome for cohorts ages 6 and 7 at the time of Head Start's launch (our control group). Column 2, which shows the ITT-event-study estimate at -1 (our estimate for cohorts that were age eligible for up to three years for a fully implemented program), suggests that Head Start significantly improved adult human capital. The standardized index increases by 1.5 percent of a standard deviation for the fully exposed cohort and 10 percent of a standard deviation for treated children (column 6). Across outcomes, column 3's ITT-spline estimates are identical to those in column 2 to the hundredth.

Supporting the impression left by the event-study plots, column 4 shows that a formal test for pre-trend (ages 6 to 15), the slope estimate for the spline component for ages 6 to 15). Figure 3.5 additionally plots the magnitude of the t-tests for the pre-trend. For the index and for each of its subcomponents, the data find no evidence of a pre-trend—a conclusion strengthened by Bonferroni-Holm (BH) p-value adjustments for multiple hypothesis testing. T-tests for pre-trends fail to reject the slope is zero in all cases. However, column 5 of Table 3.1 and Figure 3.5 shows that the data reject the null hypothesis of no trend-break at age 6 for the adult human capital index at the 1-percent level. With the exception of high school graduation (which is just below the threshold for statistical significance at the 10-percent level), the data show evidence of a trend break for each of the subcomponents of the human capital index. It appears that , even with Head Start's spill-overs to children older than 6, the relationship between adult human capital changed for children age-eligible for Head Start relative to children in the same county who were old enough to enroll in first grade.

The data show strong effects on some of the most commonly studied outcomes in the preschool literature, including both high school graduation and college enrollment. Figure 3.4b and Table 3.1 show that treated children were 1.9 percentage points more likely to complete high school/GED (column 6)—a 2.1-percent increase relative to the control mean (column 7). The magnitude of this estimate is precisely estimated, but smaller than other estimates of Head Start's effects in the literature. Figure 3.6a shows that the effect is roughly half the size of Garces, Thomas and Currie (2002)'s sibling comparison in the PSID and Thompson (2017)'s spending design in the NLSY. In addition, it is one-fifth the size of Johnson and Jackson (2017)'s spending design estimates for the very disadvantaged sample in the PSID; and one-ninth the size of Ludwig and Miller (2007)'s RD estimates using the Census. (It is one quarter the size of Deming (2009)'s sibling comparison for Head Start in the 1990s for more recent cohorts.) Although our estimate falls within the confidence intervals of previous studies, this reflects the imprecision of those estimates.

Figure 3.4c and Table 3.1 also shows a statistically significant effect of Head Start on college enrollment. Head Start raised college enrollment by 5.4 percentage points, or 8.7 percent. This estimate is half the size of Garces, Thomas and Currie (2002) and one quarter the size of Ludwig and Miller (2007) (Figure 3.6b). (The magnitude of the increase in college enrollment of 0.05 is only slightly smaller than Deming (2009)'s NLSY sibling comparison for Head Start in the 1990s.) Again, consistent with the visual impression of a trend break in Figure 3.4c, we find no evidence of a pre-trend for children older than 6 at launch and reject the null hypothesis of no trend break at age 6 at the 5-percent level.

In addition to generating more precise estimates for these commonly studied outcomes, our large-scale data permit a novel evaluation of the effects of Head Start on other dimensions of human capital, including college completion or higher degrees, which previous data have not been able to detect. Table 3.1 shows that participating children achieved 19 percent higher college graduation rates (the trend break is statistically significant at the 1-percent level). These estimates are one-quarter to one-fifth the size of those found for the Abecedarian Project (Currie, 2001; Barnett and Masse, 2007; Duncan and Magnuson, 2013). Similarly, completion of professional or doctoral degrees increased by 50 percent among treated children, although evidence of a significant pre-trend caution against a causal interpretation (this is also consistent with Figure 3.4f). These gains across the education distribution are summarized in a 0.29-year increase in schooling. This estimate is smaller than Johnson and Jackson (2017)'s estimate of 0.52 years for very disadvantaged children, but it is highly statistically significant and is not driven by a pre-trend.

These large effects on college and higher degrees may be surprising, given that no other study of preschool has documented effects on post-secondary education. This lack of evidence may reflect, in part, the small longitudinal samples or the small scale of model preschool programs. Differences in the participating children may also matter. Abecedarian and Perry's participants were very disadvantaged children and mostly black, and Perry's participants had low IQs. In contrast, Head Start was not exclusively for poor, African-American, or low-IQ children. Consequently, Head Start's participants in the 1960s and 1970s likely faced fewer socio-economic and cognitive disadvantages and less racism relative to model programs. Differences in the background characteristics of Head Start's participants make it less surprising that they experienced gains in post-secondary education.

Because analyses of model preschool programs have found different educational effects for boys, Table 3.2 stratifies our sample by sex. Among participating men, the human capital index increased by a statistically significant 14 percent of a standard deviation. For this group, high school completion rose by a statistically insignificant 2.7 percent, college attendance rose by 13 percent, and college completion rose by 27 percent. The high school estimates are smaller than others in the literature, but the college attendance estimates tend to be larger. Head Start cumulatively raised years of education among treated men by 0.41 years and the likelihood of completing a professional/doctoral degree by 59 percent. The evidence suggests that men treated with Head Start were 19 percent more likely to hold professional jobs.

The human capital index increased by less among women, at only 7 percent of a standard deviation. Completion of high school (or a GED) rose by a statistically insignificant 1.5 percent, and college attendance rose by 5.7 percent (although the trend-break is not statistically significant). For women, changes in the human capital index appear driven by increases in higher degrees, including an 11-percent increase college completion and 36-percent increase in professional degrees. Treated women's schooling rose by 0.17 years and their likelihood of holding a professional job rose by 9.5 percent.

Our Appendix Tables B.7-B.8 report estimates of Head Start's effects on human capital by race. Unfortunately, even our large sample size is too small to precisely quantify effects by race, because less than 1/6 of our sample is nonwhite and, unlike longitudinal data, we have no background covariates to model the many other determinants of adult outcomes. The broad patterns in these estimates suggest that Head Start's effects are largest among white men (13 percent of a standard deviation) and smaller among white and non-white women (5-6 percent of a standard deviation, respectively). Effects for non-white men were generally small and imprecise.

3.7 Head Start's Effects on Economic Self-Sufficiency

The substantial effects of Head Start on human capital suggest a potential for effects on economic self-sufficiency. Figure 3.7a plots the event study estimates for the self-sufficiency index, and Table 3.3 shows that an index of economic self-sufficiency aggregated over both sexes was by 4 percent of a standard deviation higher than for children ages 6-7 at the time Head Start began. Consistent with Head Start affecting less skilled individuals, the program decreased the likelihood of adult poverty by 12 percent and receipt of public assistance income by 29 percent, though these results are imprecise. Figure 3.7b and 3.7c show striking evidence of a trend break at age 6, which is also reflected in column 5 of Table 3.3 and in Figure 3.6b. However, there is little effect of Head Start on labor-force participation or wage income. (We omit the event study estimates for these outcomes for parsimony, because they are noisy and show no effect.) Null effects for wages and labor-force participation also affect the magnitude of the change and the trend-break in the economic self-sufficiency index, which does not exhibit the sharp trend-break at age 6 as in the human capital outcomes. This result may reflect the fact that men's and women's work effort changed in offsetting ways, resulting in selection for both groups. Whereas Head Start's effect on men's human capital may have led them to increase employment (e.g., the substitution effect dominates), the reverse may be true for women (e.g., the income effect dominates as more education allows them to marry higher-earning men).

Table 3.4 stratifies our sample by sex and provides evidence consistent with this hypothesis. Because the self-sufficiency estimates are noisier and stratifying by sex reduces sample sizes, we focus our discussion on the spline estimates. For treated men, the self-sufficiency index increased by 3 percent of a standard deviation. We also find positive effects of Head Start exposure on both the extensive and intensive margins of men's labor-force participation. Treated men were 2.1 percent more likely to have worked for pay (column 7), worked an average of one more week and one more hour per week (column 6). Consistent with these estimates reflecting the causal effect of Head Start, we find no evidence of a pre-trend and a marginally significant trend-break at age 6, which does not survive the Bonferroni-Holm standard error correction. At first glance, it is curious that the combined effects of increased human capital and labor-force participation do not appear to have affected annual wages. Upon further investigation, this appears consistent with Head Start inducing negative selection into the labor-force: the marginal participants tended to be less skilled, and therefore lowered the cohort's wages on average. Head Start had little effect on men's poverty, but the program is associated with a 27-percent decline in public assistance receipt among treated men. Reductions in public assistance are also consistent with negative selection, because male public assistance recipients receive high rates of disability income.

The pattern is different for women. The self-sufficiency index increased by 4 percent of a standard deviation among women treated with Head Start, largely driven by a 28-percent reduction in public assistance receipt and a 16-percent reduction in poverty. However, women's labor-force participation on the extensive and intensive margins fell slightly, albeit not significantly. These reductions in work appear to have increased annual wages of working women by around 4 percent, which is consistent with Head Start inducing positive selection (e.g., less-skilled women opting out). In our conclusion, we examine the effects of negative selection among men and positive selection among women.

As with the human capital outcomes, our Appendix Tables B.10-B.11 report estimates of Head Start's effects on self-sufficiency by race and sex. Similarly, the effects for nonwhites are generally statistically insignificant, owing to the fact that nonwhites comprise less than 15 percent of the sample and that we have few background characteristics to model the considerable variation in outcomes. As with human capital, the patterns of these estimates suggest that Head Start's effects on self-sufficiency are largest among white men and women (3 and 4 percent of a standard deviation, respectively).

3.8 Heterogeneity in Head Start's Long-Run Effect

This final section seeks to shed light on the potential mechanisms for Head Start's effects by examining how the estimates vary with access to other public programs and local economic conditions. We implement this analysis by interacting a binary indicator for whether cohorts lived in counties with "high" or "low" exposure to a program with the spline components in equation (2), where "high" is equal to one for counties above the median in the characteristics and 0 otherwise. We use only the main indices for human capital and economic self-sufficiency as the dependent variables. We caution that the lack of randomization of alternative programs and conditions means that, while these relationships are suggestive, they should not be interpreted as causal. Additionally, uncertainty about how program enrollment varied means these estimates are less precise than desired. We nevertheless provide additional guidance about the magnitudes of the ATET effects using our estimates of differential take-up across locations to scale our ITT estimates.

We first investigate the hypothesis that Head Start's long-run effects relate to their complementarities with other health programs for disadvantaged children. If health screening and referrals to health services (a sizable share of Head Start's budget) played a role in driving long-term effects, we would expect Head Start's effects to be larger for children with greater access to these health services through community health centers (CHCs) and/or Medicaid. (We would also expect the selection effects on wages to mirror those we observe for men in Table 3.4.) Table 3.5 provides suggestive evidence of this mechanism, showing that the ATETs of Head Start for human capital were slightly larger in areas with CHCs and three times as large in states where more children were eligible for Medicaid (17 percent increase relative to 5 percent for less-exposed children). The 95-percent confidence interval in the difference between these two treatment effects on the treated suggests we can reject the null hypothesis of equal effects. Consistent with health services bringing more previously disabled workers into the labor market, Head Start's ATETs on economic self-sufficiency are more muted in locations with greater health services.

A related hypothesis is that Head Start affected adult outcomes by providing healthy meals and snacks, improving child nutrition which increased both health and learning. If nutrition is an important mechanism for Head Start's long-run effects, we would expect the program's effects to be smaller for children with greater access to Food Stamps, which also supported the provision of healthy meals. Consistent with this, Table 3.5 shows that children participating in Head Start with more access to Food Stamps experienced smaller—although still statistically significant—gains in human capital and economic self-sufficiency. This suggests that the Food Stamps program provided a partial substitute for Head Start's nutritional component.

The OEO's larger effort to set up Head Start programs in the poorest 300 counties could also lead the Head Start program to be more intensive in these areas, potentially having larger effects. For this test, we report the poorest 300 counties in the column for "above median" and report the effects for counties outside this group in the column "below median." While Table 3.5 shows little evidence of differential effects on human capital across these two groups of counties, there is suggestive evidence that children in poorer counties benefitted more in terms of their economic self-sufficiency—although this difference is imprecise.

A final hypothesis is that Head Start's effects should be larger in areas with greater subsequent economic growth. Strong economic conditions should both increase the resources of children's parents, the provision of public goods (such as schools), and create stronger incentives for children to invest in themselves, as children could expect higher and more certain returns. Rather than using actual economic growth (that may be endogenous), we use predicted economic growth between 1965 and 1985. Table 3.5 suggests that the benefits of a strong economy complement Head Start's effects. The ATETs of Head Start for human capital were twice as large in areas with strong predicted economic growth than in areas with weaker predicted economic growth. The ATETs of Head Start for economic self-sufficiency were fifty percent larger large in areas with strong predicted economic growth.

All in all, these results suggest that Head Start's long-run effects may be driven by many factors beyond a preschool curriculum, including health screenings and referrals and more nutritious meals for an a population thatwith otherwise may have been under-nourished and had little access to health care and under-nourished. Unsurprisingly, the effects of Head Start appear to be complementary to the family and public resources arising from a stronger economy.

3.9 New Evidence on the Long-Term Returns to Head Start

Over the past 20 years, substantial evidence has accumulated that model preschool programs have sizable economic returns (Almond and Currie, 2011; Duncan and Magnuson, 2013; Heckman et al., 2010; Cunha and Heckman, 2007). However, convincing evidence on the long-run returns to larger-scale, public preschool has remained sparse (Phillips et al., 2017).

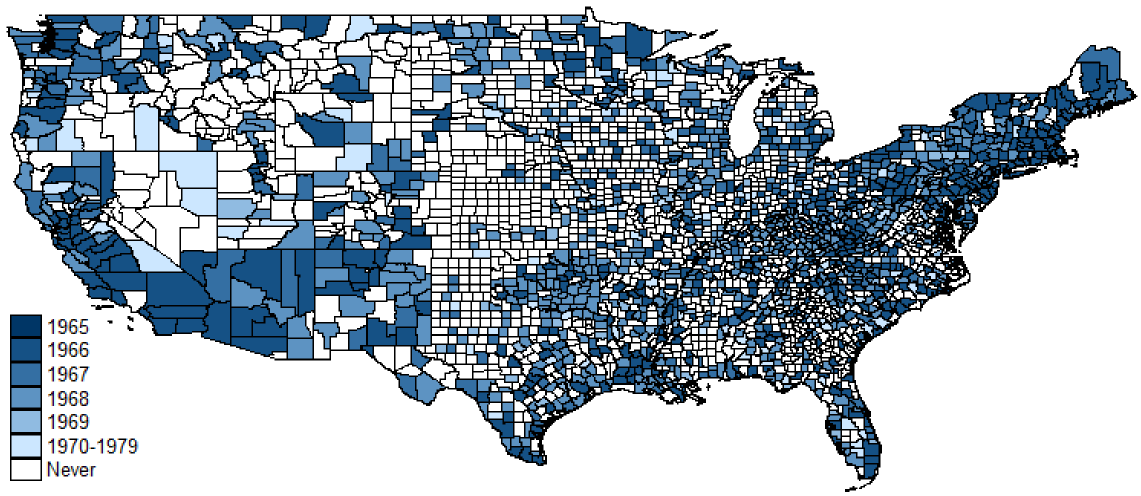
Using large-scale restricted Census/ACS data, this paper provides new evidence of the long-term effects of Head Start, the nation's longest-running, large-scale public preschool program. We find that Head Start had large effects on participants' human capital. Head Start children achieved 0.29 more years of schooling, reflecting the fact that they were 2.1 percent more likely to complete high school, 8.7 percent more likely to enroll in college, and 19 percent more likely to complete college. A second finding is that Head Start increased adult self-sufficiency, reducing the likelihood of adult poverty by 12 percent and public assistance receipt by 29 percent. Heterogeneity tests suggest that these long-run effects may reflect many aspects of the Head Start program beyond its curriculum: health screenings and referrals and more nutritious meals appear to be important mechanisms for the program's effects on disadvantaged children. In addition, the effects of Head Start appear to be complement greater family and public resources arising from a stronger economy.

A full accounting of the costs and benefits of Head Start is beyond the scope of this paper, but we summarize the implications of our estimates using potential earnings to account for selection (Neal and Johnson, 1996; Deming, 2009). Following Neal and Johnson (1996) and Deming (2009), the advantage of this approach is that it allows us to account for the effects of Head Start on employment (which differed for women and men). Like Deming, we use the NLSY79 to predict wages for individuals born from 1957 to 1965 (ages 14 to 22 in 1979)—a time frame that overlaps our Census/ACS analysis. The NLSY data allow us to estimate the relationship between wages and components of the human capital and economic self-sufficiency indices after flexibly controlling for ability using the AFQT. Although AFQT is not available in the Census/ACS, using this as a covariate helps mitigate omitted variables bias in ability in the education and earnings relationship. These regressions are reported in our Appendix Table B.6. After accounting for ability, the NLSY79 suggest a private internal rate of return to Head Start of 7.7 percent, which ranges from 4 percent for women to 11 percent for men. As an alternative, the internal rate of return of putting one child through Head Start is 2.4 percent using only savings on public assistance expenditures (estimated at $9,967 in the Survey of Income and Program Participation).

Several reasons suggest that these estimates are conservative. First, our research design differences out sibling spill-over effects, which tends to reduce the estimated effect sizes. Second, reports of income and public assistance receipt may be severely underreported in major national surveys (Meyer, Mok and Sullivan, 2015; Bound, Brown and Mathiowetz, 2001), suggesting estimates of Head Start's effect on public assistance may be understated. Third, adding increases in tax revenues and, reductions in deadweight loss from public assistance transfers, or underreporting in public assistance income would serve to increase our estimates of the returns to Head Start. Finally, estimates of the returns to Head Start ignore benefits through improvements in outcomes not measured here. For instance, they ignore the extent to which more education engenders better health, longevity, or well-being. These potential limitations, however, tend to strengthen the conclusion that Head Start achieved its goal of reducing adult poverty, delivering sizable returns to investments made in the 1960s and 1970s. The results suggest potentially larger social returns.

The long-run returns to today's public preschool programs may be different for a number of reasons. Today, the curriculum is different, the target population is different, and the alternative programs and resources available to poor children are radically different than in the 1960s. Of course, researchers will need to wait another 50 years to evaluate the long-run effects of today's preschool programs. In the meantime, the sizable returns to the "less-than-model" Head Start preschool program of the 1960s suggest productive avenues for improving the lot of poor children today.

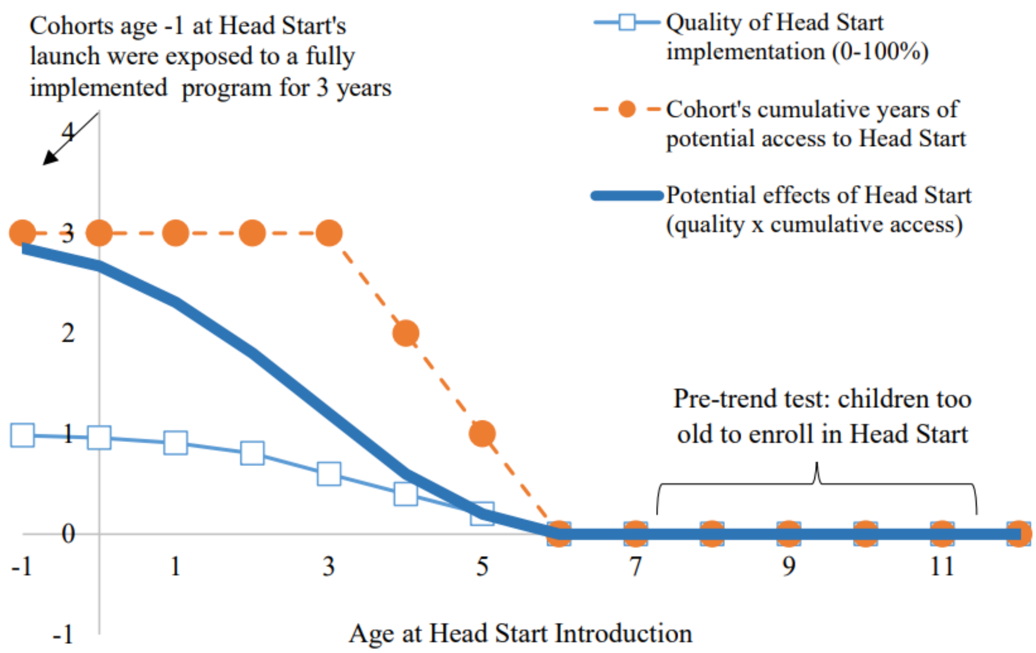
Figure 3.1: The Launch of Head Start Between 1965 and 1980



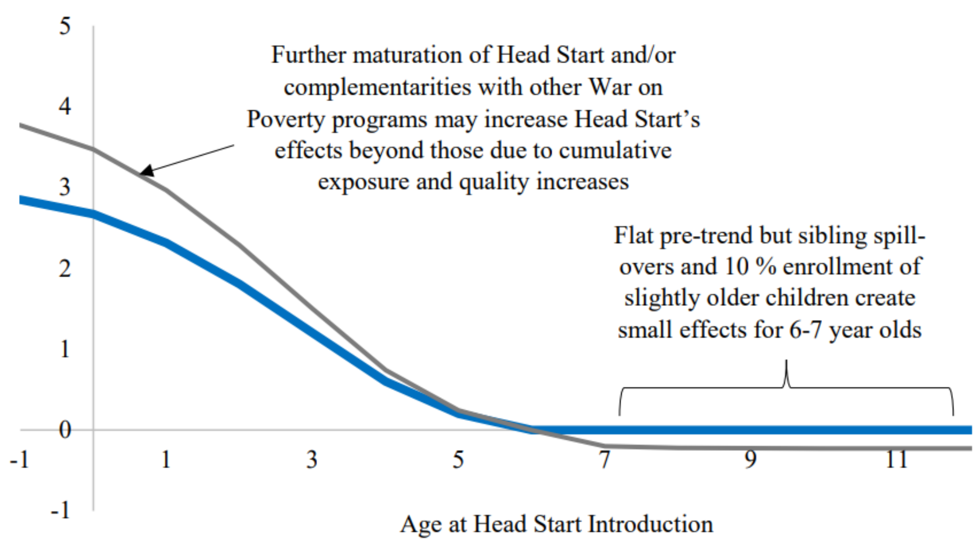
Notes: Counties are grouped by the fiscal year that Head Start launched between 1965 and 1980. Data on federal grants are drawn from the National Archives and Records Administration (NARA). See Bailey and Duquette (2014) and Bailey and Goodman-Bacon (2015) for details on data and variable construction.

Figure 3.2: The Expected Pattern of Effects on Adult Outcomes by Age of Child at Head Start's Launch

(a) No Sibling Spillovers or Complementarities with Other Programs

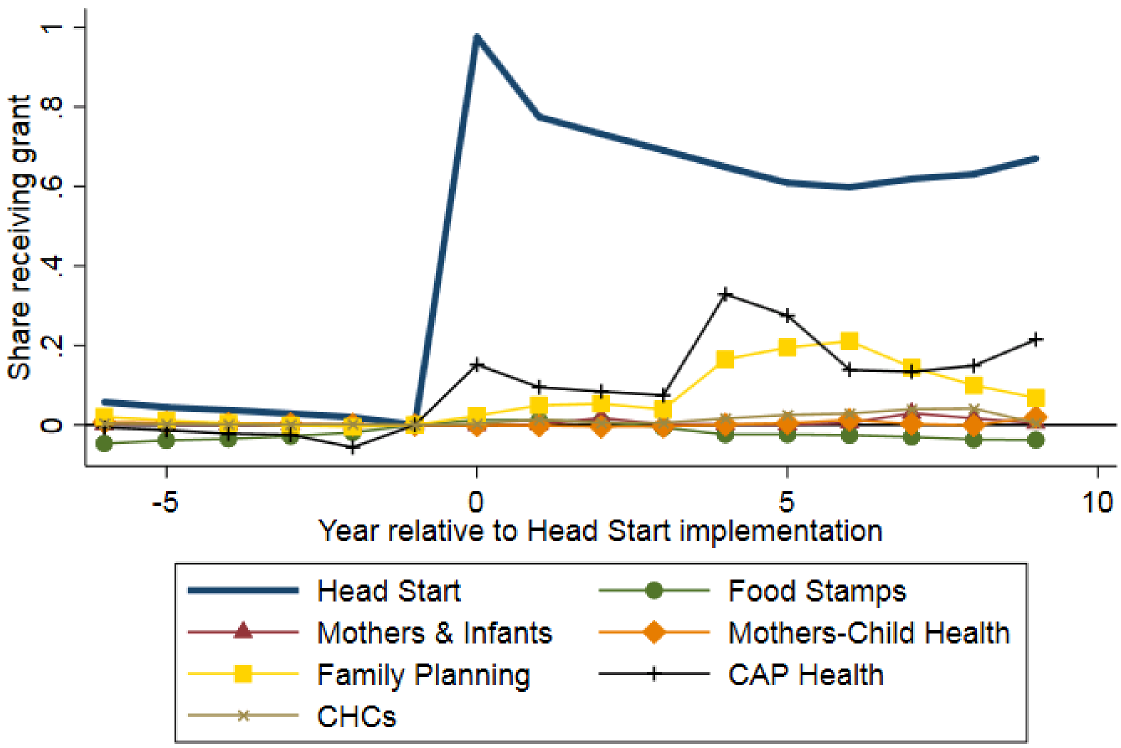


(b) With Spillovers to Siblings over 6 and Complementarities with Other Programs



Note: Figure 3.2a illustrates the potential effects of Head Start assuming there are no effects on children 6 and over, no spillovers to older siblings, and no complementarities with other programs. Figure 3.2b illustrates the way spillovers and complementarities could alter the pattern of the program's effects.

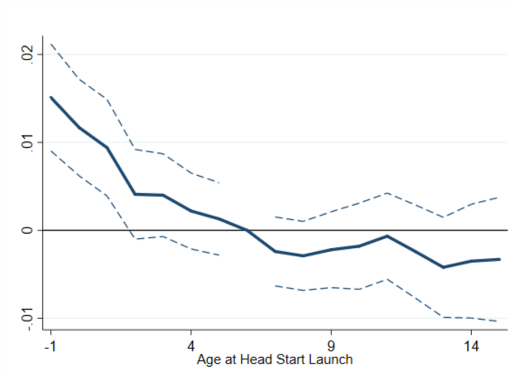
Figure 3.3: Funding for Other OEO Programs Relative to the Year Head Start Began



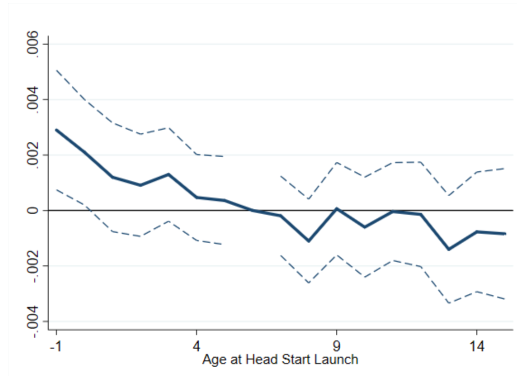
Notes: Dependent variable are binary variables for whether a county received a grant for the indicated program in the indicated year. Data on federal grants and programs are drawn from the NARA.

Figure 3.4: The Effect of Head Start on Adult Human Capital

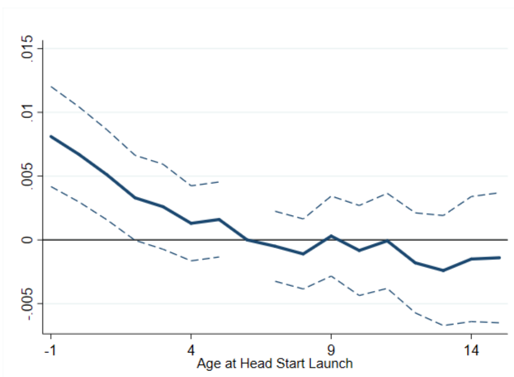
(a) Human Capital Index



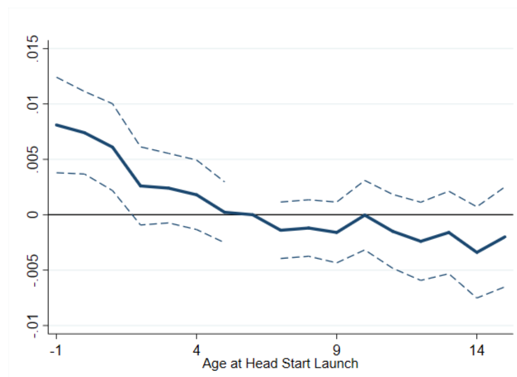
(b) High School or More



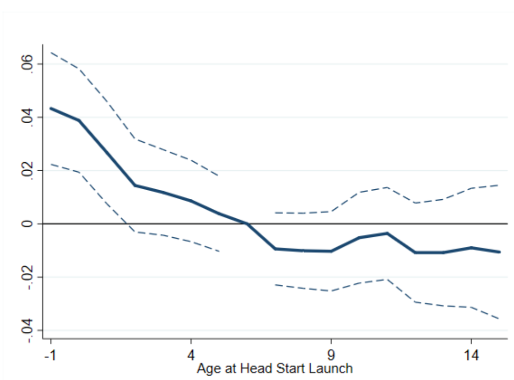
(c) Some College or More



(d) College or More



(e) Years of Schooling

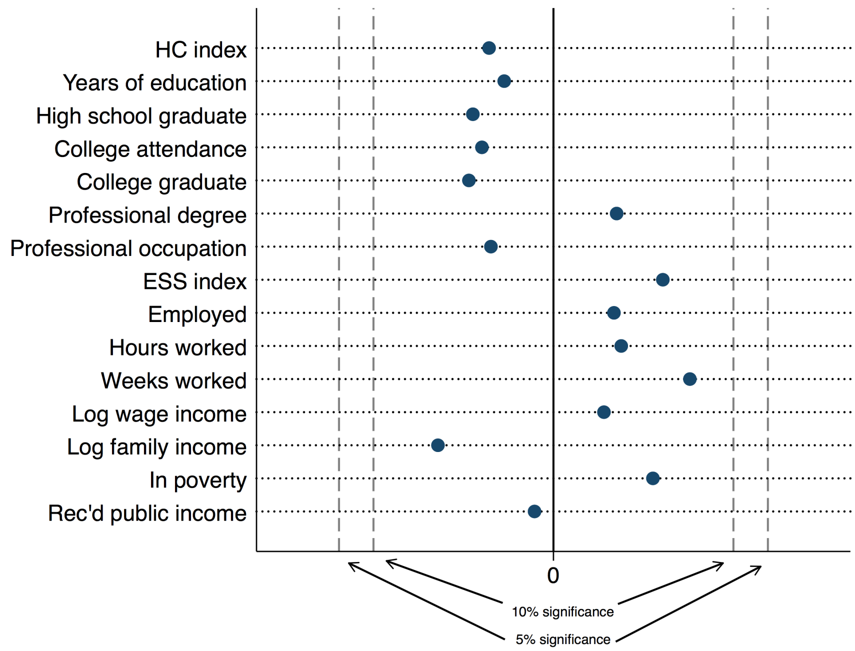


(f) Professional Job

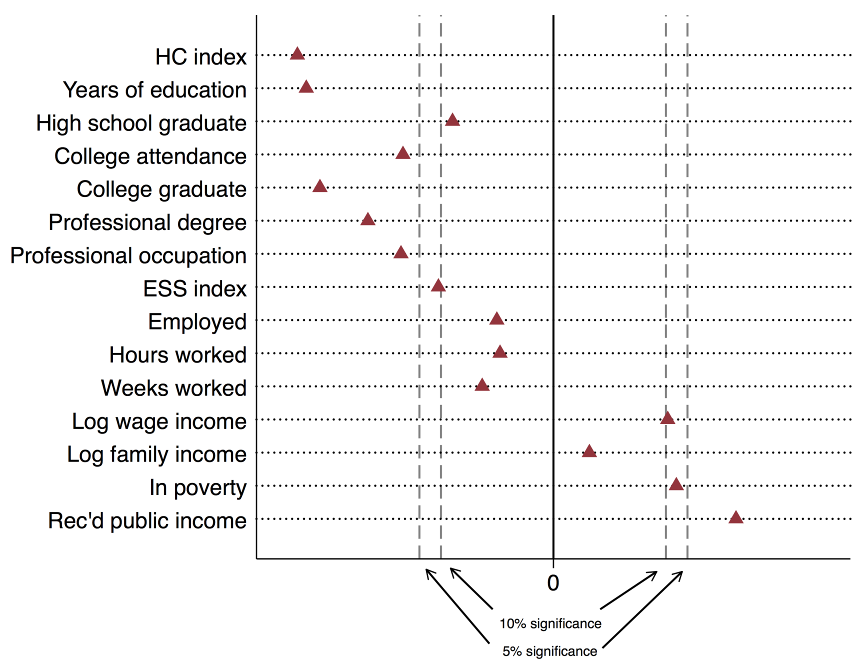
Notes: The figures plot event-study estimates of ϕ for different outcomes using the specification in equation 3.1. Standard errors clustered at the county level. Dashed lines show 95-percent, point-wise confidence intervals for each estimate.

Figure 3.5: Visual Representation of Test Statistics Evaluation Pre-Trends and Trend Breaks

(a) Test for Pre-Trend (slope of spline for age 6-15 at Head Start's launch)



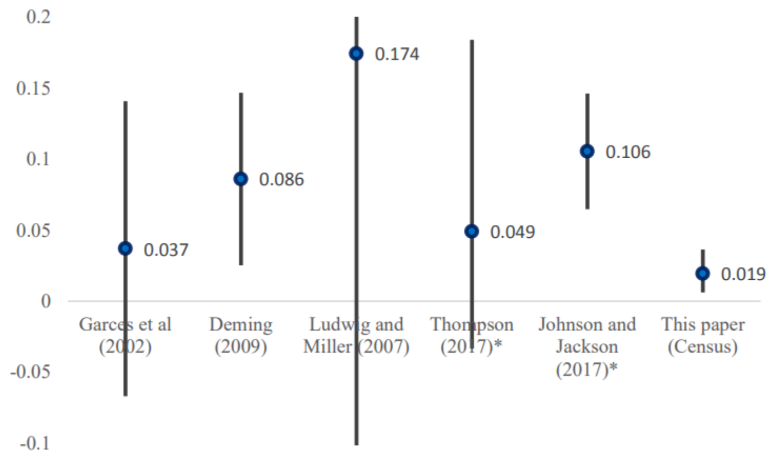
(b) Test for Trend Break (change in spline slope at age 6, before which children are age-eligible for Head Start)



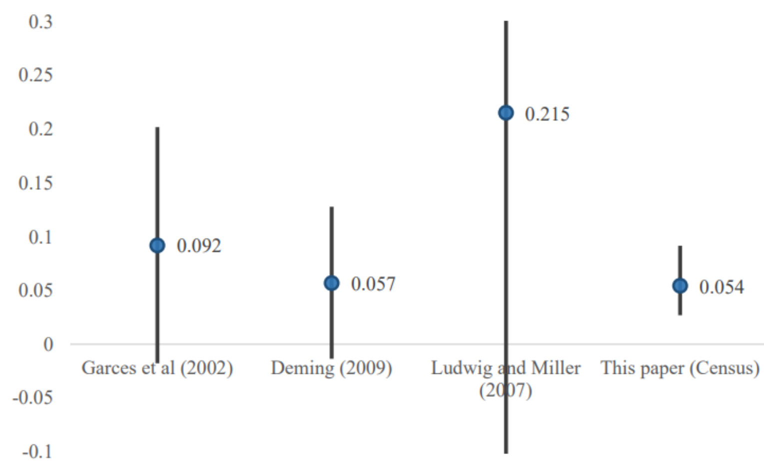
Notes: The figure plots the t-statistic on the slope of the spline for ages 6-15 (panel A) or the F-statistic for the test for a trendbreak at age 6 (panel B). Dashed lines show the threshold for statistical significance at the 10 and 5 percent levels. Compare these to columns 4 and 5 of Tables 3.1 and 3.3.

Figure 3.6: The Magnitude of Head Start's Effects on Education Across Studies

(a) Effects of Head Start on High School Graduation



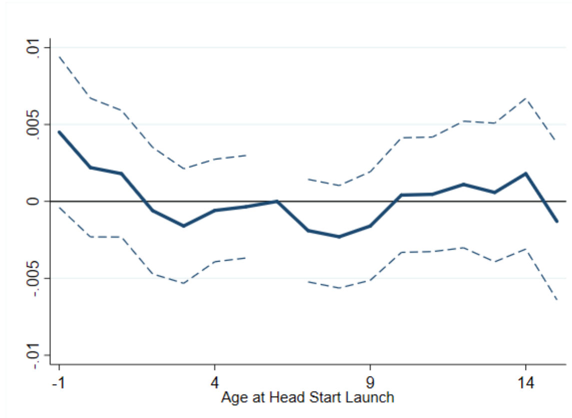
(b) The Effects of Head Start on College Enrollment



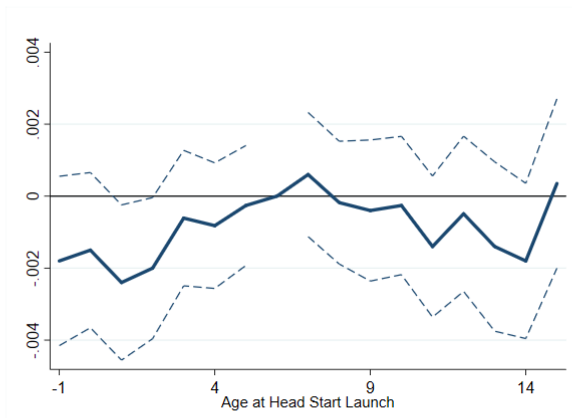
Notes: Circles indicate the reported or derived ATET from different studies. For sibling fixed effect studies, the ATET is directly reported in the papers. Because we cannot resample from data used in other Head Start papers, we calculate the ATETs for other papers using a parametric bootstrap procedure using 10,000 independent draws from normal distributions with means and standard deviations equal to the point estimates and standard errors from the reduced-form and first-stage estimates (Efron and Tibshirani, 1993). Because Johnson and Jackson (2017) does not report a standard error on the first stage, the confidence interval reported for this study in Panel A does not include this first-stage uncertainty. We limited the y-axis range so that the confidence intervals for most studies could be read from the figure. The confidence intervals for Ludwig and Miller (2007) fall outside the y-axis range and are [-0.54,1.47] in panel A and [-0.67,1.82] in panel B. Bars indicate the reported 95-percent confidence interval for sibling fixed-effect models or constructed for the ITT studies as described in the text. See Appendix for more details on the exact figures used. *Johnson and Jackson (2017) and Thompson (2017) sample likely eligible samples of the PSID and NLSY79: individuals born to parents in the bottom quartile of the income distribution, and parents with no college education, respectively.

Figure 3.7: The Effect of Head Start on Adult Economic Self-Sufficiency

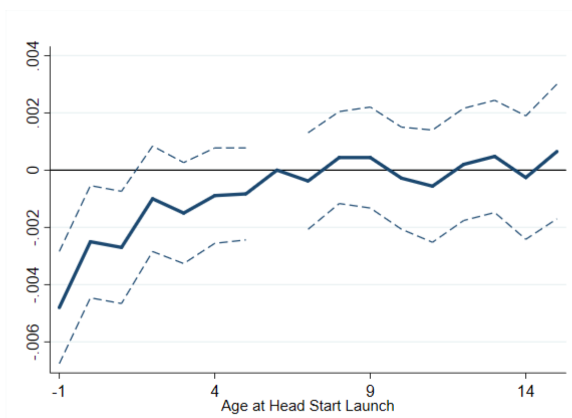
(a) Economic Self-Sufficiency Index



(b) In Poverty



(c) In Poverty



Notes: See Figure 3.4 notes. Note that In Poverty and Received Public Assistance are reverse coded when included in the Economic Self-Sufficiency Index.

Table 3.1: The Effect of Head Start on Adult Human Capital

Dependent Variables	(1) Control Mean (std.)	(2) Event Study at -1 (s.e.) [BH p-val]	(3) Spline at -1 (s.e.) [BH p-val]	(4) Test for pre-trend (s.e.) [BH p-val]	(5) F-test of trend break at age 6 (p-val) [BH p-val]	(6) ATET [95% CI]	(7) ATET % increase
Human capital index	0.014	0.015	0.015	-0.0002	14.02	0.101	
	(0.198)	(0.003)	(0.003)	(0.0003)	(0.00)	[.057,.163]	
Completed high school/GED	0.92	0.003	0.003	-0.0001	2.18	0.0189	2.1%
	(0.078)	(0.001)	(0.001)	(0.0001)	(0.14)	[.005,.038]	
		[0.01]	[0.01]	[1.00]	[0.14]		
Attended some college	0.62	0.008	0.008	-0.0002	4.84	0.054	8.7%
	(0.140)	(0.002)	(0.002)	(0.0003)	(0.03)	[.027,.092]	
		[0.00]	[0.00]	[1.00]	[0.08]		
Completed 4 year college	0.29	0.008	0.009	-0.0002	11.66	0.054	18.6%
	(0.122)	(0.002)	(0.002)	(0.0002)	(0.00)	[.025,.094]	
		[0.00]	[0.00]	[1.00]	[0.00]		
Prof. or doc. degree	0.028	0.002	0.002	0.0000	7.36	0.014	50.0%
	(0.037)	(0.001)	(0.000)	(0.0001)	(0.01)	[.006,.024]	
		[0.00]	[0.00]	[1.00]	[0.03]		
Years of schooling	13.57	0.043	0.049	-0.0005	13.06	0.291	2.1%
	(0.695)	(0.011)	(0.010)	(0.0012)	(0.00)	[.144,.49]	
		[0.00]	[0.00]	[1.00]	[0.00]		
Has a professional job	0.35	0.007	0.007	-0.0001	4.97	0.0489	14.0%
	(0.121)	(0.002)	(0.002)	(0.0002)	(0.03)	[.022,.085]	
		[0.00]	[0.00]	[1.00]	[0.08]		

Notes: In column 1, the control mean and standard deviation are calculated using the cohorts ages 6 and 7 at the time Head Start was launched. Column 2 presents the estimated intention-to-treat (ITT) effect evaluated at birth cohort of full exposure (-1, see Figure 3.2). Column 3 presents the ITT spline estimate evaluated at -1. Column 4 presents the pre-trend estimate for the spline segment for age 6 and older at implementation. Column 5 presents the F-statistic and p-value for the test of a trend-break in the spline at age 6. The ATET estimate in column 6 divides the ITT effect at -1 by the estimate of receiving a Head Start grant on school enrollment at school age 5, 0.149 (s.e. 0.022) for the full sample and 0.151 (s.e. 0.022) for men and 0.145 (s.e. 0.022) for women; see Appendix Table A5). Column 7 computes the percentage increase implied by the ATET relative to the control mean (the ratio of Column 6 to Column 1) for components of the index. The BH p-values presented in columns 2, 4, and 5 in brackets use the Bonferroni-Holm method to account for multiple hypotheses testing of individual outcomes within an index.

Table 3.2: The Effect of Head Start on Adult Human Capital by Sex

Dependent Variables	(1) Control Mean (std.)	(2) Event Study at -1 (s.e.) [BH p-val]	(3) Spline at -1 (s.e.) [BH p-val]	(4) Test for pre-trend (s.e.) [BH p-val]	(5) F-test of trend break at age 6 (p-value) [BH p-val]	(6) ATET [95% CI]	(7) ATET % increase
A. Men							
Human capital index	0.014	0.021	0.02	-0.0005	11.79	0.1360	
	(0.247)	(0.004)	(0.003)	(0.0004)	(0.00)	[.081,.215]	
Completed high school/GED	0.91	0.004	0.003	-0.0001	1.92	0.0250	2.7%
	(0.104)	(0.002)	(0.001)	(0.0002)	(0.17)	[.005,.049]	
		[0.01]	[0.01]	[1.00]	[0.22]		
Attended some college	0.59	0.011	0.01	0.0000	5.67	0.0740	12.5%
	(0.176)	(0.003)	(0.002)	(0.0003)	(0.02)	[.036,.125]	
		[0.00]	[0.00]	[1.00]	[0.07]		
Completed 4 year college	0.29	0.012	0.011	-0.0003	9.99	0.0769	26.5%
	(0.150)	(0.003)	(0.002)	(0.0003)	(0.00)	[.041,.126]	
		[0.00]	[0.00]	[1.00]	[0.01]		
Prof. or doc. degree	0.034	0.003	0.003	0.0000	4.73	0.0199	58.5%
	(0.052)	(0.001)	(0.001)	(0.0001)	(0.03)	[.008,.035]	
		[0.00]	[0.01]	[1.00]	[0.09]		
Years of schooling	13.5	0.062	0.063	-0.0012	12.15	0.4129	3.1%
	(0.878)	(0.013)	(0.011)	(0.0015)	(0.00)	[.228,.668]	
		[0.00]	[0.00]	[1.00]	[0.00]		
Has a professional job	0.34	0.01	0.009	-0.0006	2.54	0.0649	19.1%
	(0.153)	(0.003)	(0.002)	(0.0003)	(0.11)	[.031,.11]	
		[0.00]	[0.00]	[0.06]	[0.22]		
B. Women							
Human capital index	0.015	0.01	0.012	0.0000	6.50	0.0659	
	(0.228)	(0.004)	(0.003)	(0.0004)	(0.01)	[.014,.132]	
Completed high school/GED	0.93	0.002	0.002	-0.0001	1.42	0.0140	1.5%
	(0.087)	(0.001)	(0.001)	(0.0001)	(0.23)	[-.003,.035]	
		[0.28]	[0.07]	[1.00]	[0.47]		
Attended some college	0.65	0.005	0.006	-0.0004	1.20	0.0370	5.7%
	(0.163)	(0.002)	(0.002)	(0.0003)	(0.27)	[.005,.077]	
		[0.15]	[0.01]	[0.97]	[0.47]		
Completed 4 year college	0.29	0.005	0.008	-0.0002	4.63	0.0320	11.0%
	(0.147)	(0.003)	(0.002)	(0.0003)	(0.03)	[-.004,.076]	
		[0.28]	[0.01]	[1.00]	[0.16]		
Prof. or doc. degree	0.022	0.001	0.001	0.0001	3.78	0.0080	36.4%
	(0.042)	(0.001)	(0.001)	(0.0001)	(0.05)	[-.001,.018]	
		[0.28]	[0.07]	[0.91]	[0.16]		
Years of schooling	13.6	0.024	0.038	-0.0003	4.91	0.1659	1.2%
	(0.795)	(0.013)	(0.012)	(0.0014)	(0.03)	[-.013,.379]	
		[0.28]	[0.01]	[1.00]	[0.16]		
Has a professional job	0.37	0.005	0.005	0.0004	4.35	0.0350	9.5%
	(0.150)	(0.003)	(0.002)	(0.0003)	(0.04)	[0,.078]	
		[0.25]	[0.07]	[0.97]	[0.16]		

Notes: See Table 3.1 notes.

Table 3.3: The Effect of Head Start on Adult Economic Self-Sufficiency

Dependent Variables	(1) Control Mean (std.)	(2) Event Study at -1 (s.e.) [BH p-val]	(3) Spline at -1 (s.e.) [BH p-val]	(4) Test for pre-trend (s.e.) [BH p-val]	(5) F-test of trend break at age 6 [p-val] [BH p-val]	(6) ATET [95% CI]	(7) ATET % increase
Self-sufficiency index	0.024	0.0055	0.0048	0.0003	4.24	0.037	
	(0.154)	(0.0024)	(0.0022)	(0.0002)	(0.04)	[.005,.076]	
Worked last year	0.86	0.00082	0.00091	0.0000	0.22	0.006	0.7%
	(0.098)	(0.0014)	(0.0013)	(0.0002)	(0.64)	[-.013,.026]	
		[1.00]	[1.00]	[1.00]	[1.00]		
Weeks worked last year	41.1	0.047	0.055	0.0057	0.61	0.32	0.8%
	(5.281)	(0.075)	(0.0721)	(0.0092)	(0.43)	[-.68,1.4]	
		[1.00]	[1.00]	[1.00]	[1.00]		
Usual hours works per week	35.72	0.0334	0.056	0.0106	1.09	0.22	0.6%
	(4.907)	(0.080)	(0.0791)	(0.0085)	(0.30)	[-.84,1.4]	
		[1.00]	[1.00]	[1.00]	[1.00]		
Log labor income	10.6	0.0055	0.0061	0.00018	2.8	0.037	
	(0.268)	(0.0035)	(0.0031)	(0.0004)	(0.09)	[-.009,.09]	
		[0.70]	[0.30]	[1.00]	[0.47]		
Log family income/poverty	5.86	0.0059	0.0054	-0.0004	0.28	0.04	
	(0.258)	(0.0043)	(0.0041)	(0.0004)	(0.60)	[-.017,.104]	
		[0.70]	[0.75]	[1.00]	[1.00]		
In poverty*	0.1	-0.0018	-0.002	0.0001	3.23	-0.012	-12.0%
	(0.084)	(0.0012)	(0.0011)	(0.0001)	(0.07)	[-.03,0.004]	
		[0.70]	[0.35]	[1.00]	[0.44]		
Rec'd public assistance*	0.11	-0.0048	-0.0037	0.0000	7.13	-0.032	-29.1%
	(0.093)	(0.0010)	(0.0008)	(0.0001)	(0.01)	[-.052,-.018]	
		[0.00]	[0.00]	[1.00]	[0.05]		

Notes: *In poverty and received public program income are reverse-coded when used in the self-sufficiency index. See also Table 3.1 notes

Table 3.4: The Effect of Head Start on Adult Economic Self-Sufficiency by Sex

Dependent Variables	(1) Control Mean (std.)	(2) Event Study at -1 (s.e.) [BH p-val]	(3) Spline at -1 (s.e.) [BH p-val]	(4) Test for pre-trend (s.e.) [BH p-val]	(5) F-test of trend break at age 6 [p-val] [BH p-val]	(6) ATET [95% CI]	(7) ATET % increase
A. Men							
Self-sufficiency index	0.031	0.0046	0.00553	-0.0001	4.24	0.0299	
	(0.201)	(0.0031)	(0.0025)	(0.0003)	(0.21)	[-.01,.077]	
Worked last year	0.91	0.0029	0.00301	0.0000	2.33	0.0189	2.1%
	(0.108)	(0.0014)	(0.0011)	(0.0001)	(0.13)	[.001,.041]	
		[0.23]	[0.04]	[1.00]	[0.58]		
Weeks worked last year	44.79	0.153	0.168	-0.0015	2.46	1.013	2.3%
	(6.031)	(0.0780)	(0.0637)	(0.0078)	(0.12)	[.002,2.221]	
		[0.25]	[0.04]	[1.00]	[0.58]		
Usual hours works per week	41.31	0.148	0.1722	0.0039	2.94	0.98	2.4%
	(6.040)	(0.0882)	(0.0798)	(0.0080)	(0.09)	[-.161,2.317]	
		[0.37]	[0.12]	[1.00]	[0.52]		
Log labor income	10.88	0.0033	0.00238	-0.0008	0.33	0.0219	
	(0.301)	(0.0043)	(0.0036)	(0.0005)	(0.57)	[-.035,.083]	
		[1.00]	[1.00]	[0.61]	[1.00]		
Log family income/poverty	5.91	0.0035	0.00245	-0.0006	0.07	0.023	
	(0.302)	(0.0048)	(0.0041)	(0.0005)	(0.79)	[-.04,.092]	
		[1.00]	[1.00]	[1.00]	[1.00]		
In poverty*	0.07	-0.00033	-0.000637	-0.0001	0.01	-0.002	-2.9%
	(0.095)	(0.0013)	(0.0011)	(0.0001)	(0.90)	[-.02,.016]	
		[1.00]	[1.00]	[1.00]	[1.00]		
Rec'd public assistance*	0.11	-0.0046	-0.0046	0.0002	9.93	-0.029	-26.4%
	(0.110)	(0.0015)	(0.0011)	(0.0001)	(0.00)	[-.056,-.01]	
		[0.02]	[0.00]	[1.00]	[0.01]		
B. Women							
Self-sufficiency index	0.019	0.0058	0.00406	0.0005	4.24	0.0399	
	(0.18)	0.0029	0.0027	0.0004	[0.10]	[.001,.088]	
Worked last year	0.8	-0.0018	-0.00168	0.0001	0.15	-0.012	-1.5%
	(0.13)	(0.0020)	(0.0018)	(0.0003)	(0.70)	[-.043,.016]	
		[0.90]	[1.00]	[1.00]	[1.00]		
Weeks worked last year	37.66	-0.097	-0.0931	0.0091	0.03	-0.667	-1.8%
	(6.80)	(0.1100)	(0.1015)	(0.0146)	(0.87)	[-2.31,.858]	
		[0.90]	[1.00]	[1.00]	[1.00]		
Usual hours works per week	30.49	-0.11	-0.092	0.0117	0.00	-0.745	-2.4%
	(5.81)	(0.0970)	(0.0950)	(0.0129)	(0.95)	[-2.24,.599]	
		[0.90]	[1.00]	[1.00]	[1.00]		
Log labor income	10.31	0.0062	0.007	0.00084	3.94	0.043	
	(0.34)	(0.0051)	(0.0048)	(0.0005)	(0.04)	[-.026,.122]	
		[0.90]	[0.57]	[0.61]	[0.28]		
Log family income/poverty	5.81	0.0083	0.0084	-0.00034	0.69	0.057	
	(0.32)	(0.0054)	(0.0050)	(0.0005)	(0.41)	[-.015,.143]	
		[0.62]	[0.48]	[1.00]	[1.00]		
In poverty*	0.12	-0.0028	-0.00273	-0.0003	4.21	-0.018	-15.0%
	(0.11)	(0.0017)	(0.0015)	(0.0002)	(0.04)	[-.047,.004]	
		[0.60]	[0.38]	[0.71]	[0.28]		
Rec'd public assistance*	0.12	-0.0049	-0.0029	0.0002	0.67	-0.034	-28.3%
	(0.12)	(0.0014)	(0.0011)	(0.0001)	(0.41)	[-.06,-.014]	
		[0.00]	[0.06]	[0.71]	[1.00]		

Notes: *In poverty and received public program income are reverse-coded when used in the self-sufficiency index. See also Table 3.1 notes

Table 3.5: Heterogeneity in the Effect of Head Start, by Local Programs and Economic Circumstances

	Intent-to-Treat (ITT) Effect			Effect on preschool enrollment (s.e.)		Average Treatment Effect on Treated Children (ATET)	
	Above median	Below median	F-test of diff (p-val)	Above median	Below median	Above median	Below median
A. Human capital index							
Medicaid exposure	0.021	0.009	4.095	0.123	0.175	0.171	0.052
	(0.004)	(0.004)	(0.043)	(0.024)	(0.030)		
CHC exposure	0.020	0.010	9.258	0.203	0.109	0.100	0.090
	(0.003)	(0.003)	(0.002)	(0.034)	(0.025)		
Food Stamps exposure	0.011	0.020	4.515	0.122	0.167	0.092	0.117
	(0.003)	(0.004)	(0.034)	(0.026)	(0.032)		
Poorest 300th counties	0.009	0.015	0.320	0.073	0.125	0.125	0.123
	(0.011)	(0.003)	(0.572)	(0.026)	(0.023)		
Predicted economic growth	0.023	0.010	11.948	0.139	0.125	0.166	0.078
	(0.004)	(0.003)	(0.001)	(0.033)	(0.025)		
B. Self-sufficiency index							
Medicaid exposure	0.003	0.005	0.336	0.123	0.175	0.023	0.031
	(0.004)	(0.002)	(0.562)	(0.024)	(0.030)		
CHC exposure	0.006	0.003	1.708	0.203	0.109	0.030	0.028
	(0.003)	(0.003)	(0.191)	(0.034)	(0.025)		
Food Stamps exposure	0.002	0.007	3.547	0.122	0.167	0.017	0.039
	(0.003)	(0.002)	(0.060)	(0.026)	(0.032)		
Poorest 300th counties	0.004	0.004	0.000	0.073	0.125	0.060	0.035
	(0.010)	(0.002)	(0.991)	(0.026)	(0.023)		
Predicted economic growth	0.006	0.003	1.014	0.139	0.125	0.041	0.027
	(0.003)	(0.003)	(0.314)	(0.033)	(0.025)		

Notes: ATETs are constructed by dividing the group-specific ITT estimate of Head Start's effect on long-run outcomes by the group-specific estimated first stage. Results for the 300 poorest counties are reported in the column for "above median" with results for other counties reported in "below median."

APPENDICES

APPENDIX A

Appendix to The Long-Run Effects of America's First Paid Maternity Leave Policy

A.1 STDI coverage and the implementation of anti-pregnancy discrimination laws

My identification strategy relies on two sources of variation that interacted to create a staggered, state-level expansion of paid maternity leave in the United States. First, a series of states, and eventually the federal government, enacted anti-discrimination laws that required short-term disability insurance (STDI) to cover childbirth as a disability. These laws effectively created a source of paid maternity leave benefits for women covered by STDI, and the differential timing of their enactment allows me to compare outcomes of women and children within states and over time in an event-study specification. The second source of variation comes from long-standing differences in *access* to short-term disability insurance, driven largely by state disability policies and industrial mix. This second source of variation meant that the enactment of anti-discrimination laws had more "bite" in some states than in others.

To my knowledge, there exists no comprehensive history of anti-pregnancy-discrimination laws in the United States. Nor is there a comprehensive source of data on STDI coverage at a sub-national level.

To assemble evidence on the enactment of state anti-discrimination laws and the receipt of STDI benefits, I rely on several primary and secondary sources, including Congressional testimony, correspondence with state officials, newpaper articles, and published histories of anti-discrimination laws (Gladstone, Williams and Belous, 1985; Kamerman, Kahn and Kingston, 1983; U.S. Senate, 1979; U.S. House of Representatives, 1977). These laws varied in their specifics and in the way they were enacted. While the policy was enacted in some states by legislative action, others were created by a ruling through the state Supreme Court or an action of the executive branch of government. In addition, for those states that did not enact anti-discrimination laws before 1979, the policy was imposed on them by the U.S. Congress through the Pregnancy Discrimination Act of 1978. The resulting timeline is laid out in Table A.1.

These anti-discrimination laws also varied in their scope. Many affected a very broad range of workers. In the case of the Pregnancy Discrimination Act, the ban on pregnancy discrimination affected firms with 15 or more employees. In states where STDI was nearly universal, the share of affected workers would have been even larger. In addition, many of the laws, including the federal Pregnancy Discrimination Act, required only that "women affected by pregnancy, childbirth, or related medical conditions shall be treated the same for all employment-related purposes" as men or women who were not pregnant but had conditions that affected their ability to work in similar ways. In these cases, STDI maternity benefits were not the only effect of the laws. For example, Gruber (1994) explores the wage and employment effects of Pregnancy Discrimination Act-driven changes in health insurance benefits for childbirth. As a result, I cannot completely rule out the possibility that the labor-market effects I estimate aren't driven in part by other subtle changes that resulted from the anti-discrimination laws. However, a reading of the legislative history and newspaper accounts suggests that maternity leave was the dominant concern among proponents and opponents of the legislation. I specifically discuss my results in the context of Gruber (1994) below.

The second source of variation that I exploit in my research design – differences in *access* to STDI – dates to the origins of the industry in the 19th century. The goal of early STDI policies was to provide financial stability to workers, typically males, who wanted insurance against the risk of an injury or illness that would prevent them from earning income (Faulkner, 1940). The STDI industry grew significantly over the early 20th century, and throughout the second half of the 20th century, about 60 percent of workers were covered (Price, 1986). As shown in Figure A.1, this coverage rate remained steady throughout much of the 1970s and 1980s. However, the stability of aggregate STDI coverage rates belies substantial variation across states. Coverage was much more prevalent among workers in certain industries, such as manufacturing (Levy, 2004). As a result, state-level STDI coverage rates varied with the mix of industries in existence. In addition, five states – Rhode Island, New Jersey, New York, California, and Hawaii – enacted laws in the 1940s (and in the 1960s, in the case of Hawaii) that made access to STDI virtually universal. This variation in access to STDI existed decades before the anti-pregnancy-discrimination laws of the 1970s and was driven primarily by the desire for wage insurance among workers, rather than concerns about allowing women to take leave after the birth of a child.

While the Social Security Administration has tracked data on national STDI coverage levels, little information is available on the share of workers covered at the state level. To capture this variation, I construct a measure of access to STDI that relies only industry-level STDI coverage and the distribution of female workers across industries in 1970, before anti-discrimination laws were enacted in states without universal STDI coverage. This measure therefore avoids relying on endogenous responses from firms that may change STDI coverage in response to the anti-discrimination laws. The industry-level shares come from a tabulation prepared by the Bureau of Labor Statistics National Compensation Survey and published in Autor et al. (2013). The share of women employed by state and industry comes from the 1970 long-form decennial Census, accessed via IPUMS (Ruggles et al., 2017). The resulting estimates range from 26 percent to universal coverage and are shown in Figure A.2.

A.1.1 An additional test of the research design

One concern about the use of these two sources of variation to estimate the effects of paid maternity leave is that the staggered implementation of anti-discrimination laws may be systematically related to other state-level characteristics that are related to working mothers' leave-taking behavior, women's labor-market outcomes, or children's long-run educational attainment. Such cross-state differences could confound my estimates of the effect of STDI maternity benefits.

The main text of this paper discusses the key assumptions behind my identification strategy. Several tests of these assumptions provide little evidence for concern. First, the use of state fixed effects will eliminate any time-invariant confounding factors, while division-by-year fixed effects go further by netting out any differential trends common to certain regions of the United States. Differential *trends* could also present problems; however, the built-in test for pre-trends in my event-study research design suggest no evidence of such confounding factors. I also find no evidence of coincident changes in other programs, such as the Earned Income Tax Credit, that could explain the sharp break in female hourly wages and children's outcomes that I report.

Figure A.4 provides an additional test for systematic relationships between the implementation of STDI maternity benefits and state-level characteristics that could drive the changes in labor-market activity and educational attainment that I report. The figure shows the t-statistics from a regression of the year of enactment of the state anti-discrimination law on a set of 21 measures of economic and demographic characteristics from the 1960 Census. I restrict the set of characteristics to those used in a similar exercise by Bailey (2006) in an analysis of the the effect of state laws that affected access to the birth control pill. In addition, I drop New Jersey and Rhode Island from the sample because these states began paying STDI maternity benefits at least a decade before any other state.

Figure A.4 displays results that are consistent with a legislative history, as reflected in Table A.1, that suggests the implementation of STDI benefits were driven more by idiosyncratic factors than systematic differences across states. Only one of the 21 covariates delivers a t-statistic that exceeds the traditional 5% level, an outcome we would expect in an exercise that features 21 statistical tests. In addition, the lone significant result suggests that average education among women in early-adopting states was lower, providing a counterpoint to the possibility that early-adopting states were systematically driven by a more educated, empowered female electorate.

A.2 Relevance of STDI benefits

The literature on parental leave has produced voluminous evidence of the effect of maternity benefits on women's leave-taking and attachment to the workforce. However, little work has been done to quantify the effect of STDI maternity benefits on women's leave-taking. In addition, an understanding of the take-up of these benefits is crucial to the interpretation of the estimated effects on women's labor-market wages and employment and children's long-run outcomes.

My preferred estimates of the effect of STDI maternity benefits on women's leave-taking rely on the retrospective fertility module of the 1984 and 1985 panels Survey of Income and Program Participation. These estimates, which are reported in the main text of the paper, suggest that women who received benefits took an extra 5-6 weeks away from work around childbirth. In this section, I provide complementary estimates from two alternative sources of data.

A.2.1 Evidence from the decennnial Census

One limitation of my preferred estimates of take-up of STDI maternity benefits is that they rely on retrospective responses from the SIPP that are asked only of women who worked during their first pregnancy. These features of the data limit the statistical power and, potentially, the generalizability of the result.

Another source of data is the 1970 and 1980 decennial Census, accessed via IPUMS (Ruggles et al., 2017). These data report the year, quarter, and state of birth for each respondent, and allow them to be connected to parents if they reside in the same household. In addition, they include questions about employment status in the previous week and the previous year.

I construct a sample of women ages 18-45 who gave birth to a child in the calendar quarter preceding the Census reference date, which was April 1 in each Census year. For an additional comparison group, I also use a sample of women age 18-45 who report that they have never given birth. I then estimate my main difference-in-difference specification using three binary outcomes: being absent from a job in the previous week, employed in the previous week, and working for pay at any time during the previous year. The first outcome provides an estimate of the effect on leave-taking. The second provides some additional context about the effect of the policy; a rise in leave-taking accompanied by a rise in employment would suggest the paid-leave benefits increase mothers' attachment to the workforce, while a rise in leave-taking without a change in employment suggests STDI-funded leave results in a short-run substitution from time at work to time at home with no longer-run implications for mothers' labor-market attachment (Baker and Milligan, 2008). Finally, the measurement of whether mothers worked for pay in the previous calendar year provides a look at whether the availability of paid-leave benefits affected women's labor-supply before childbirth.

Table A.2 displays the results from the decennials Census, with the effects on mothers in Panel A. The probability of being on leave after enactment of STDI maternity benefits rose by 1.8 percentage points – a 42 percent increase relative to the base of 4.23 percent. However, in columns 2 and 3, I see no effect on employment and, perhaps surprisingly, a substantial negative effect on the share of women working before childbirth. Columns 4 and 5 restrict the sample to women who report working for pay the previous year. Because STDI maternity benefits are only available to women with a work history, this purges the sample of women who were not eligible for benefits and so should not respond to the availability of benefits. As we would expect, we see a much stronger increase in leave-taking and an increase in employment that is nearly as large. Taken together, the results of Panel A suggest that, as proponents of parental leave policies suggest, the availability of paid benefits increases leave-taking and makes working women more likely to remain employed after the birth of a child. However, the estimate in column 3 also suggests a deterioration in women's labor market prospects.

Panels B and C of Table A.2 provide additional robustness checks. If the employment and leave-taking effects are truly driven by the availability of maternity leave, we would not expect to see effects on women who have never given birth. Consistent with this story, the estimates in Panel B are all indistinguishable from 0. One notable exception is the estimate in column 3, the likelihood of working for pay in the previous year. This estimate is still more negative than the corresponding estimate for new mothers, such that if I use non-mothers as a comparison group in a triple-difference specification as in Panel C, I get a small positive point estimate for the probability that new mothers worked the previous year. Altogether, these results show clearly that STDI maternity benefits increased leave-taking among new mothers. They also show, consistent with evidence from the CPS, that women saw a deterioration in labor-market conditions, although the effects in the Census appear on the employment margin while the CPS shows effects only for hourly wages.

A.2.2 Evidence from the 1984-1989 SIPP

An additional source of data that can be used to evaluate the effect of STDI maternity benefits on women's short-run labor-market activity is the 1984-1989 panels of the SIPP. While sample sizes are smaller than those offered by the decennial Census, these data provide detailed information on the week-to-week labor-market activity of respondents. Relative to the retrospective data I use for my preferred estimates, they also rely on reports of labor-market activity that are more recent – and therefore, possibly less prone to error. The chief drawback is that these data go back only to 1984, several years after the last set of anti-discrimination laws were enacted, so I cannot compare women who gave birth before STDI benefits became available to those who gave birth afterward. I rely instead on within-person variation in labor-market activity and cross-state variation in access to STDI.

The results are shown graphically in Figure A.3. The left-hand panel shows that, in states without universal STDI policies, women reduce the time spent at work, relative to 1 year before birth, by nearly 50 percent in the month of birth and more than 50 percent in the month after. However, the relative decrease is even larger where paid benefits are available, falling over 60 percent relative to 1 year before birth.

Table A.3 shows the points estimates from a specification that pools the coefficients into multiple-month bins to make the results easier to digest. Column 1 displays results for all working mothers. The estimate for each time period is statistically indistinguishable from 0 except for the months closest to childbirth, where women with universal access to STDI benefits spend an additional 0.112 months – or 2-3 workdays – at home rather than the workplace. Summing over four months, this amounts to an intent-to-treat effect of 0.45 fewer months, or about 2 fewer weeks, at work around the birth of a child for the average working woman. Using my estimate of 0.4 as a take-up rate for STDI benefits among eligible women, this translates to an increase of 1.125 months of leave – or close to 5 weeks of leave. This estimate is comparable to my preferred estimate using retrospective data.

Columns 2 through 5 suggest this effect is relatively constant across race and education groups. Relative to their counterparts in non-universal-STDI states, less-educated women reduce time spent at work slightly more than women who attended college. This is consistent with an effect driven by STDI, since higher-educated women are more likely to have maternity benefits offered through a private arrangement with an employer, even in the wake of government mandates. The point estimate for non-white women in the months around childbirth is slightly lower than that for the full sample or white women, although the standard error is too large to rule out equal or even larger effects than other groups. On the whole, for women who remained in the workforce, STDI-funded maternity benefits appear to have led to longer spells of parental leave, including for the least-advantaged members of the workforce who would be less likely to receive benefits without government policy.

A.3 Maternity leave and the results of Gruber (1994)

My identification strategy is based on an expansion of paid maternity leave via STDI, which was required to cover childbirth as a disability as a result of the Pregnancy Discrimination Act of 1978 and a number of state-level precursors. A closely related paper is Gruber (1994), which examined the effect of some of these same policies on the wages and employment of married men and women. Unlike this paper, Gruber (1994) focuses on another consequences of these anti-discrimination laws: Health insurance policies were required to cover the hospital charges of women who give birth.

The Pregnancy Discrimination Act did not explicitly conditions that STDI or health insurance plans were required to cover. Rather, it stated only that firms must treat women who cannot work before, after, or during the birth of a child the same way they would treat any other employee who is temporarily unable to work. One consequence of this broadly worded policy is that it is ultimately not possible to separate labor-market effects that are driven by maternity leave from those driven by health insurance or other factors.

However, several pieces of evidence suggest that maternity leave benefits were indeed one of the most significant consequences of these anti-discrimination laws, and that the results of Gruber (1994) may be worth reinterpreting accordingly.

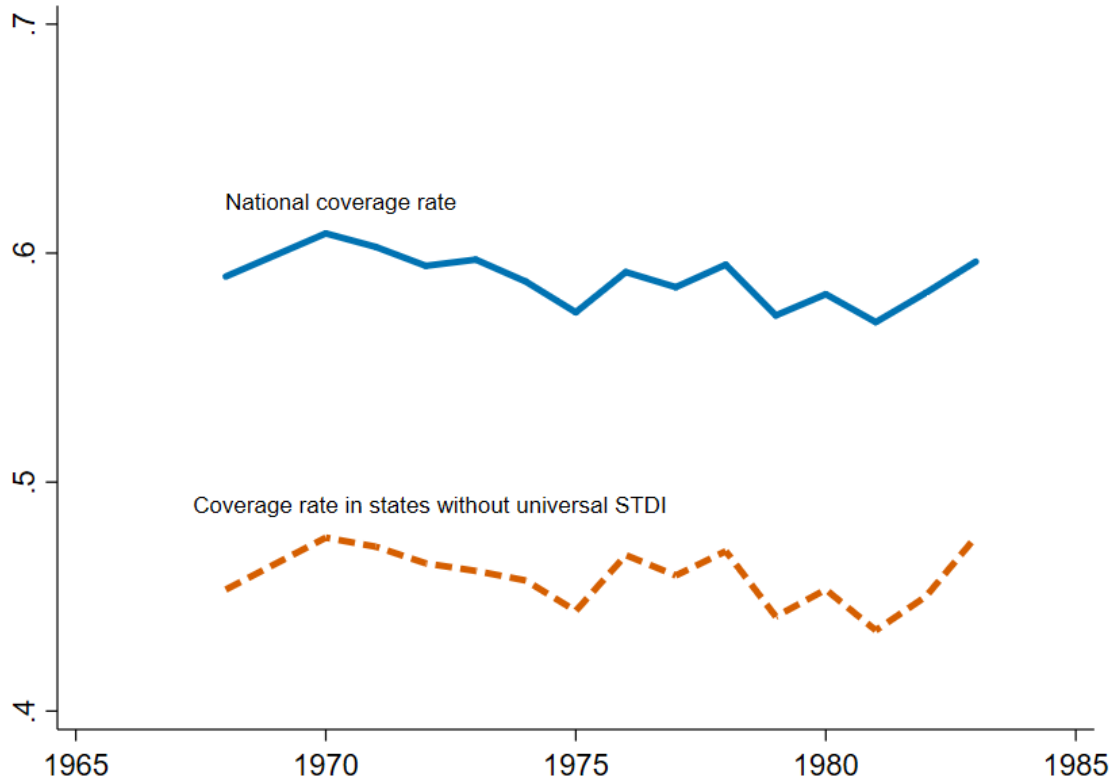
The first set of evidence worth noting is the qualitative evidence from debates in Congress and statehouses over the Pregnancy Discrimination Act and its state-level counterparts. Maternity benefits through STDI were a primary objection from business groups opposing the legislation, who argued that it would not only raise STDI premiums but would also lead to longer and more frequent leaves that would force firms to hire less productive, temporary workers and increase turnover.

Additional evidence can be gleaned from a useful feature of the state-level variation used to estimate the main wage effects in Gruber (1994). In particular, some of the main results use a triple-difference strategy that, in part, compares married women from three early-adopting states – Illinois, New York, and New Jersey – to women from a set of control states that enacted anti-discrimination laws later. In two of these three states, STDI maternity benefits and health insurance maternity benefits were enacted at the same time. However, in New Jersey, STDI maternity benefits were enacted much earlier, in 1961. If the observed effects on wages were driven by factors other than STDI maternity benefits, we would expect to see strong wage responses in all three states if we estimated the effects separately. However, if STDI maternity benefits are the most salient consequence of the anti-discrimination laws, then New Jersey should react quite differently than the other states.

In fact, the evidence from a replication of the findings of Gruber (1994) suggests that STDI maternity benefits were indeed the major driver of labor-market responses to the anti-discrimination laws of thoe 1970s. These results are shown in Table A.4. Columns 1 and 2 show the main result from Table 4 of Gruber (1994) and my replication, respectively. In column 3, I alter the specification by replace the indicator for treated states (referred to as "experimental" states in the paper) with a binary indicator for each treatment state, allowing me to estimate the wage effect separately for each state. The results show that while New York and Illinois saw large negative wage effects in the wake of the passage of their respective anti-discrimination laws, New Jersey saw virtually no impact. This suggests that the anti-discrimination laws were actually heterogeneous, ushering in STDI maternity benefits in New York and Illinois but imposing much smaller costs on New Jersey, where maternity benefits had been available for more than a decade.

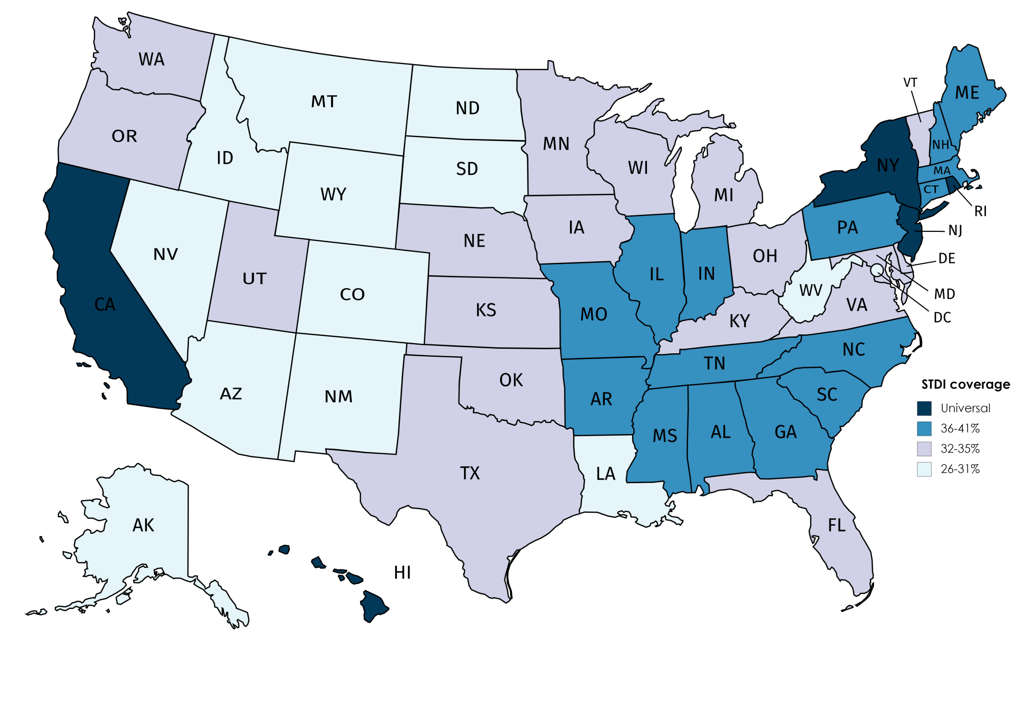
A.4 Figures and Tables

Figure A.1: National share of workers covered by STDI



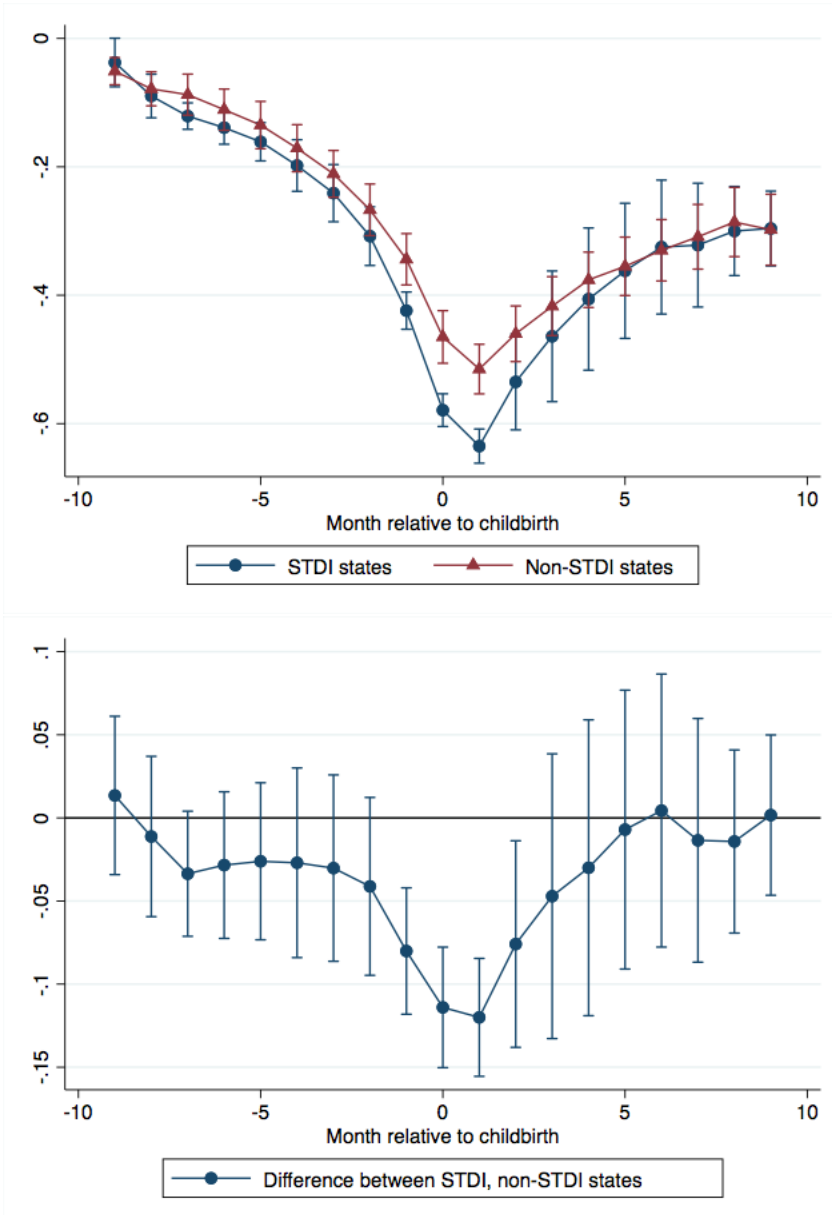
Notes: Data obtained from Price (1986) and shows share of workers covered by STDI in all states (solid line) and in states without universal STDI coverage (dashed line). States with universal STDI coverage are California, New Jersey, New York, and Rhode Island, plus Hawaii beginning in 1969.

Figure A.2: Estimated share of working women with STDI coverage in 1970



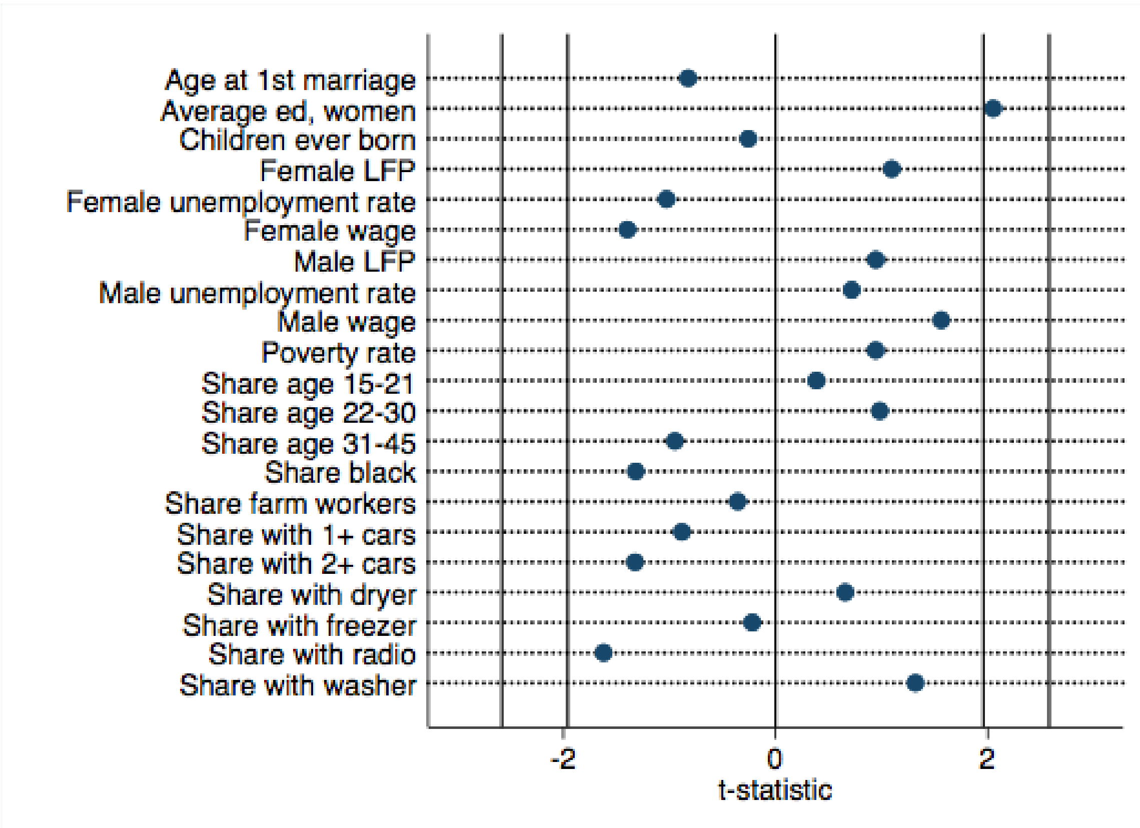
Notes: State-level estimates are constructed using industry-wide STDI coverage shares from the Bureau of Labor Statistics' National Compensation Survey, as published in Autor et al. (2013), and the share of women employed by industry and state from the 1970 decennial Census accessed via IPUMS (Ruggles et al., 2017). Dark blue states are those with state laws requiring near-universal STDI coverage.

Figure A.3: Effect of STDI on time spent at work by month



Data: Survey of Income and Program Participation, 1984-1989 panels. Sample is limited to women who give birth during a SIPP panel and who were employed 12 months before childbirth. Left-hand panel shows regression-adjusted mean of share of month spent at work relative to 12 months before birth. Right-hand panel shows the difference between STDI and non-STDI states.

Figure A.4: Correlation of anti-discrimination law passage and state characteristics



Notes: Plot shows t-statistics from multivariate regression with dependent variable of year STDI-funded maternity leave benefits were enacted at the state level. Regressions are weighted by the 1960 state population. Data on state characteristics comes from the 1960 long-form decennial Census accessed via IPUMS (Ruggles et al., 2017).

Table A.1: Variation in timing and intensity of the expansion of STDI pregnancy benefits

	Universal STDI adopted	Pregnancy benefits adopted	Mode of passage
Rhode Island	1942	1942	Legislature
New Jersey	1948	1961	Legislature
Montana	–	1972	Legislature
Connecticut	–	1973	Legislature
Hawaii	1969	1973	Legislature
Alaska	–	1975	Legislature
Iowa	–	1975	State court
Kansas	–	1975	Administrative
South Dakota	–	1975	Administrative
Wisconsin	–	1975	State court
Illinois	–	1976	Administrative
California	1946	1977	Legislature
Maryland	–	1977	Legislature
Michigan	–	1977	Legislature
New York	1949	1977	State court
Washington, DC	–	1977	Legislature
Massachusetts	–	1978	State court
All other states	–	1979	Congress

Notes: Column 1 lists date that state adopted universal STDI law, where applicable. For all non-universal states, between 26 and 41 percent of working women had coverage (see Figure A.2). Column 2 shows the year each state's anti-pregnancy-discrimination law was enacted. Column 3 lists the political entity that spurred enactment of the anti-pregnancy-discrimination law.

Table A.2: Intent-to-treat effects on working mothers

	Full sample			DI-eligible	
	(1)	(2)	(3)	(4)	(5)
	On leave	Employed	DI-eligible	On leave	Employed
Panel A: New mothers					
STDI x Post	0.0179**	-0.00182	-0.0417***	0.0540**	0.0508***
	(0.00852)	(0.0154)	(0.00979)	(0.0206)	(0.0172)
	[0.194]	[0.937]	[0.021]	[0.116]	[0.093]
Panel B: Women without children					
STDI x Post	-0.00719	-0.0380	-0.0581**	-0.00860	-0.00351
	(0.00527)	(0.0385)	(0.0258)	(0.00714)	(0.0262)
	[0.416]	[0.481]	[0.156]	[0.505]	[0.948]
Panel C: Triple-difference					
STDI x Post x new mother	0.0253**	0.0388	0.0201	0.0640***	0.0532**
	(0.0103)	(0.0287)	(0.0325)	(0.0172)	(0.0217)
	[0.108]	[0.397]	[0.665]	[0.048]	[0.142]
Mean: Moms in STDI states, 1970	0.0423	0.113	0.486	0.0851	0.217
Total observations	1,099,109	1,099,109	1,099,109	917,511	917,511

Notes: Coefficients displayed are estimated intent-to-treat effects of exposure to paid maternity benefits on employment and leave-taking, from equation (1) in panels A and B and equation (2) in panel C. Variable D_s is the imputed share of working women with access to STDI benefits, constructed as described in the text. Coefficient vector X_{istg} includes a quadratic in age. Sample includes treatment group of women age 18-45 who gave birth in the last three months, and control group of women of same age with no children, from long-form 1970 and 1980 decennial Census via IPUMS (Ruggles et al., 2017). Standard errors in parentheses are clustered at the state level. Figures in brackets are p-values from a two-sided permutation test of the null hypothesis of no effect of paid maternity leave.

Table A.3: Effect on working mothers: Complementary evidence from the SIPP

	(1)	(2)	(3)	(4)	(5)
	All working mothers	High school or less	Some college	White	Nonwhite
Dependent variable: Share of month at work					
6-9 months before childbirth	-0.00984	-0.00721	-0.00615	-0.0185	0.0699
	(0.0288)	(0.0450)	(0.0252)	(0.0247)	(0.0441)
2-5 months before	-0.0252	-0.0143	-0.0291	-0.0256	-0.0621
	(0.0335)	(0.0501)	(0.0247)	(0.0299)	(0.0958)
1 month before to 2 months after	-0.112***	-0.112**	-0.0952***	-0.109***	-0.0791
	(0.0263)	(0.0423)	(0.0261)	(0.0223)	(0.0858)
3-6 months after	-0.0734	-0.0259	-0.0770	-0.0636	0.111
	(0.0480)	(0.0422)	(0.0604)	(0.0558)	(0.106)
7-10 months after	-0.0388	0.0113	-0.0630	-0.0375	0.0544
	(0.0407)	(0.0467)	(0.0618)	(0.0471)	(0.108)
Observations	41,108	21,096	18,276	36,080	5,256
R squared	0.501	0.507	0.504	0.505	0.505
Mean in STDI states	0.820	0.776	0.870	0.830	0.758
Mean in other states	0.811	0.731	0.880	0.807	0.831

Notes: Coefficients displayed are estimated intent-to-treat effects of exposure to paid maternity benefits on the share of each month spent at work from a modified version of equation (3) that pools $\boldsymbol{\tau}$ into specified groups of months relative to childbirth. Sample means are weighted averages of share of month at work 10 or more months before childbirth. Sample includes women from the 1984-1989 panels of the Survey of Income and Program Participation who give birth during a panel and are aged 18-45 at the time of childbirth.

Table A.4: Replication of Gruber (1994): Effect on log wages

	(1) Gruber 1994	(2) Replication	(3) Replication, by state
Treatment x Post x Experimental State	-0.043*	-0.0467**	
	(0.023)	(0.0224)	
Treatment x Post x Illinois			-0.0627**
			(0.0311)
Treatment x Post x New York			-0.0594**
			(0.0288)
Treatment x Post x New Jersey			-0.00692
			(0.0382)
Observations	27,033	26,971	26,971
R-squared		0.414	0.415

Notes: Table shows replication of results of Gruber (1994) Table 4. Sample includes married women ages 20-40 ("treatment" group) and single men age 20-40 from the 1974, 1975, 1977, and 1978 May CPS. Specification is based on equation 1 in Gruber (1994) and includes controls for years of education, quadratic in potential experience, full interaction of gender and marital status, indicators for nonwhite race and union membership, and year fixed effects. In column 3, equation 1 has been modified to report β_8 separately for each "experimental" state.

APPENDIX B

Appendix to The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency

B.1 Census Data

The project's primary data source is the 2000 Census and 2001 to 2013 ACS combined with the SSA's Numident file, accessed through project 1284 in the University of Michigan Research Data Center (RDC). The advantage of these data is that they link a rich set of productivity outcomes for cohorts potentially benefitting from War on Poverty programs (those who are ages 25 to 54 in the Census/ACS) with information on their access to Head Start programs in childhood using Numident information on county of birth. The 2000 census long-form contains information on 16.7 percent of the U.S. population; the 2001 to 2013 ACSs contain information for around 14 percent of the U.S. population. The number of Numidentlinked, unique individuals in these combined data sources represent about one-quarter of the U.S. population. In addition, we use the 1970 restricted long-form Census that contains information on school enrollment for children. Unfortunately, these data cannot be linked to the Numident because they have not yet been PIK'd by the Census.

B.2 SSA's Numident: Data on County and Date of Birth

Links from the Census/ACS to place and date of birth are important for studying the long-term impacts of Head Start, as place and date of birth provide crucial information on exposure of individuals to these programs in early childhood. For the Census/ACS files we use the survey-internal PIK code to match individuals with the SSA's Numident file. The Numident place-of-birth variable is a string variable detailing, in most cases, the city and state of birth. In previous work, Isen et al. (2013) developed a matching algorithm to connect this string variable to the Census Bureau database of places, counties, and minor civil divisions as well as the United States Geological Survey's Geographic Names Information System (GNIS) file. We also make use of code that was developed for a similar purpose by Black et al. (2015). Using both sources, we constructed a crosswalk between the NUMIDENT place of birth string variable and (standard) county FIPS codes, with over 90 percent of individuals matched to their counties of birth. Taylor et al. (2016)'s technical memorandum has been posted with the Census Bureau and contains this information.

The Census/ACS data have the benefit of providing a wide range of outcomes of interest, including individual earnings, but also program participation, disability, living arrangements, and family and household variables such as income and poverty. Additionally, we observe individuals in all states and we observe individuals regardless of whether they are employed. The Census/ACS data also have limitations. They are repeated cross-sections and information is self-reported (and, so, measured with error).

B.3 Data on School Age Entry Cutoffs

We restrict our sample to individuals who were born in areas where we have information for the relevant school-entry age cut-offs. This information is taken from Bedard and Dhuey (2012) and supplemented using our own research for two states. According to state legislative documents, the school-entry age cutoff for the entire state of Texas was September 1st starting in 1969.[1] The state of Kansas' cutoff date is January 1st before 1965.[2]

We omit areas from our analysis sample where we are missing information on school entry cutoffs. This includes all individuals born in Colorado, Georgia, Indiana, Massachusetts, Montana before 1979, New Jersey, Rhode Island before 1967, South Carolina before 1978, Texas before 1969, Utah, Washington before 1977, and West Virginia before 1972. In our robustness checks, we find that the inclusion of these states and cohorts using a generic school-entry age cutoff tends to attenuate the estimates, as one might expect in the case of classical measurement error.

B.4 Data on Head Start

To study the long-run impacts of access to Head Start, we also use additional county-level sources of data to account for potentially confounding local programs and the economy. These data include information from the following sources: Bailey and Goodman-Bacon (2015) collected data on the OEO's community programs from the National Archives Community Action Program (NACAP) files, as well as from some administrative sources. For Community Health Centers, some information was hand-entered from annual Public Health Service (PHS) Reports. The resulting database contains information on (1) the county where a program delivered services, which allows each federal grant to be linked to county-level mortality rates; (2) the date that each county received its first program services grant (this excludes planning grants), which provides the year that programs began operating; and (3) some information on program grants between 1978 and 1980 from the National Archives Federal Outlays (NAFO) files. We supplement these data with information on the legal services program from Cunningham (2013) and Food Stamps from Hoynes and Schanzenbach (2009).

[1] http://www.heinonline.org/HOL/Page?men_tab=srchresults&handle=hein.ssl/sstx0163&size=2&collection=ssl&id=718

[2] http://www.heinonline.org/HOL/Page?men_tab=srchresults&handle=hein.ssl/ssks0074&size=2&collection=ssl&id=477

B.5 Data on Head Start Launch Dates

Our main policy variable is the availability of the Head Start program, which was rolled out across counties during the War on Poverty. Bailey and Duquette (2014) and Bailey and Goodman-Bacon (2015) have compiled information from the National Archives and Records Administration on changes in Head Start funding between 1965 and 1980, which we use in this study. To verify their accuracy, these data have been compared to federal government directories of Head Start programs (Project Head Start 1971, Office of Child Development 1973, Project Head Start 1978). The following tables and figures supplement the analysis and description in the paper with more information on the roll-out of the program.

Table B.1: Share of Counties and Children under 6 in County with Head Start, 1965-1980

Fiscal Year HS Launched	# of Counties	Percent of counties	Cumulative percent of counties	Percent of kids age under 6 in 1970	Cumulative percent of kids age under 6 in 1970
1966	536	17.53%	17.57%	54.71%	55.22%
1967	217	7.10%	24.66%	10.57%	65.79%
1968	607	19.86%	44.52%	16.06%	81.85%
1969	41	1.34%	45.86%	0.71%	82.56%
1970	45	1.47%	47.33%	0.87%	83.43%
1971	10	0.33%	47.66%	0.17%	83.61%
1972	30	0.98%	48.64%	0.51%	84.12%
1973	7	0.23%	48.87%	0.11%	84.23%
1974	9	0.29%	49.17%	0.15%	84.38%
1975	16	0.52%	49.69%	0.29%	84.67%
1976	7	0.23%	49.92%	0.08%	84.75%
1977	4	0.13%	50.05%	0.16%	84.90%
1978	9	0.29%	50.34%	0.31%	85.21%
1979	3	0.10%	50.44%	0.08%	85.29%
No HS<1980	1515	49.56%	100.00%	14.71%	100.00%
Total	3057				

Table B.2: 1960 County Characteristics and Head Start's Launch, 1965-1980

	(1)	(2)	(3)	(4)	(5)
	Head Start launches before 1980	No Head Start before 1980	First Head Start grant in		
			1965-1966	1967-1968	1969-1980
# of Counties	1542	1515	537	824	181
County population	87,460	15,326	161,162	52,056	29,971
% Urban	72.21	31.02	79.44	61.4	42.35
% Rural Farm	5.72	24.44	3.42	9.01	16.38
% Nonwhite	10.77	11.2	11.33	9.45	12.31
% Population Aged 0-4	11.53	11.15	11.51	11.56	11.58
% Population aged 65+	8.95	10.89	8.74	9.29	9.66
Median Family Income (1959)	5712.77	4145.36	5984.64	5311.17	4550.95
Active MD per 1,000 Pop	0.69	0.02	0.96	0.23	0.04
AMR, All Ages	955.45	925.51	963.51	941.67	935.72
Infant Mortality Rate	25.6	27.28	25.36	25.77	28.01
Any Med Students (1969)	0.25	0.01	0.35	0.1	0
% Family Income <$3000	20.41	35.85	18.02	23.85	31.39
% Family Income $10,000+	15.15	7.44	16.73	12.74	9.04
% w/ 12+ Yrs Schooling (Age 25+)	41.81	34.18	43.04	40.01	36.37
% w/ <4 Yrs Schooling (Age 25+)	8.22	10.8	7.79	8.74	10.94
Gov'ts Exp ($000s) per 1,000 Pop (1957)	143.29	128.68	152.01	128.96	117.34
Labor Force Participation	0.38	0.36	0.38	0.37	0.36
% Labor Force Unemployed	5.32	4.8	5.38	5.21	5.24

Notes: All values are population-weighted means, with the exception of average county population in row 1. Characteristics are for 1960 unless otherwise specified. All variables are taken from the 1960 County and City Databooks (Haines et al. 2010) and 1990 Area Resource Files (US DHHS 1994) except the following. Medicare variables are for 1966, taken from the County-level Medicare File (US SSA 1969-1977; US HFA 1978-1980). Data on Medicare expenditures were shared by Almond, Hoynes and Schanzenbach (2011). We also use the 1959 to 1988 Vital Statistics Multiple-Cause of Death Files (US DHHS and NCHS 2007) to compute mortality rates.

Table B.**3:** Regression Analysis of 1960 County Characteristics and Head Start's Launch, 1965-1980

	(1)	(2)	(3)
% Nonwhite	0.008	0.006	0.005
	[0.006]	[0.007]	[0.007]
% Population Aged 0-4	0.014	0.053	0.059
	[0.060]	[0.063]	[0.061]
% Population Aged 65+	0.002	0.021	-0.001
	[0.033]	[0.035]	[0.035]
Median Family Income (1959)	0.000	0.000	0.001
	[0.000]	[0.000]	[0.000]
Active MD per 1000 Pop	-0.755***	-0.540***	0.395***
	[0.192]	[0.182]	[0.145]
AMR, All Ages	0.000	-0.000	-0.000
	[0.001]	[0.001]	[0.001]
AMR, 1960-65 Change	-0.000	0.000	0.000
	[0.001]	[0.001]	[0.001]
Infant Mortality Rate	0.000	-0.000	-0.003
	[0.005]	[0.006]	[0.006]
Any Med Students (1969)	-0.238	-0.439**	-0.401**
	[0.187]	[0.211]	[0.190]
% Family Income <$3,000	0.028	0.017	0.023
	[0.027]	[0.028]	[0.027]
% Family Income $10,000+	-0.090**	-0.090*	-0.064
	[0.042]	[0.048]	[0.046]
% w/ 12+ Yrs Schooling (Age 25+)	0.013	0.020	0.012
	[0.014]	[0.017]	[0.017]
% w/ <4 Yrs Schooling (Age 25+)	0.001	0.007	0.011
	[0.017]	[0.020]	[0.020]
Gov'ts Exp ($000s) per 1000 Pop (1957)	0.002	0.003	0.001
	[0.002]	[0.002]	[0.002]
Labor Force Participation	5.325***	5.417***	2.863
	[1.789]	[1.813]	[1.797]
% Labor Force Unemployed	-0.068**	-0.061**	-0.040
	[0.027]	[0.028]	[0.027]
% Labor Force Male	0.054***	0.050***	-0.006
	[0.017]	[0.018]	[0.019]
Constant	1,959***	1,960***	1,969***
	[3.305]	[3.621]	[3.608]
Observations	1,510	1,510	1,510
R-squared	0.065	0.113	0.157
Covariates		S,U	S,U,P

Notes: We estimate each regression by ordinary least squares. The dependent variable is year of Head Start's launch. Heteroskedasticity robust standard errors are beneath the point estimates in brackets. The regressions exclude counties that never received Head Start funding and are unweighted. Characteristics are for 1960 unless otherwise specified. Covariates without point estimate in the table include state fixed effects (S), urban categories (U, 0, $0 < u \leq 25$, $25 \leq u < 50$, $50 \leq u < 75$, $75 \leq u \leq 100$, where u is the share of a county's population living in an urban area), and log population (P). See also Table B.2 notes.

Table B.**4:** Age at Launch by Cohort and Year Head Start Launched

	Head Start launch year														
	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
Birth cohort															
1950	15														
1951	14	15													
1952	13	14	15												
1953	12	13	14	15											
1954	11	12	13	14	15										
1955	10	11	12	13	14	15									
1956	9	10	11	12	13	14	15								
1957	8	9	10	11	12	13	14	15							
1958	7	8	9	10	11	12	13	14	15						
1959	6	7	8	9	10	11	12	13	14	15					
1960	5	6	7	8	9	10	11	12	13	14	15				
1961	4	5	6	7	8	9	10	11	12	13	14	15			
1962	3	4	5	6	7	8	9	10	11	12	13	14	15		
1963	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1964	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1965	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1966	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
1967		-1	0	1	2	3	4	5	6	7	8	9	10	11	12
1968			-1	0	1	2	3	4	5	6	7	8	9	10	11
1969				-1	0	1	2	3	4	5	6	7	8	9	10
1970					-1	0	1	2	3	4	5	6	7	8	9
1971						-1	0	1	2	3	4	5	6	7	8
1972							-1	0	1	2	3	4	5	6	7
1973								-1	0	1	2	3	4	5	6
1974									-1	0	1	2	3	4	5
1975										-1	0	1	2	3	4
1976											-1	0	1	2	3
1977												-1	0	1	2
1978													-1	0	1
1979														-1	0
1980															-1

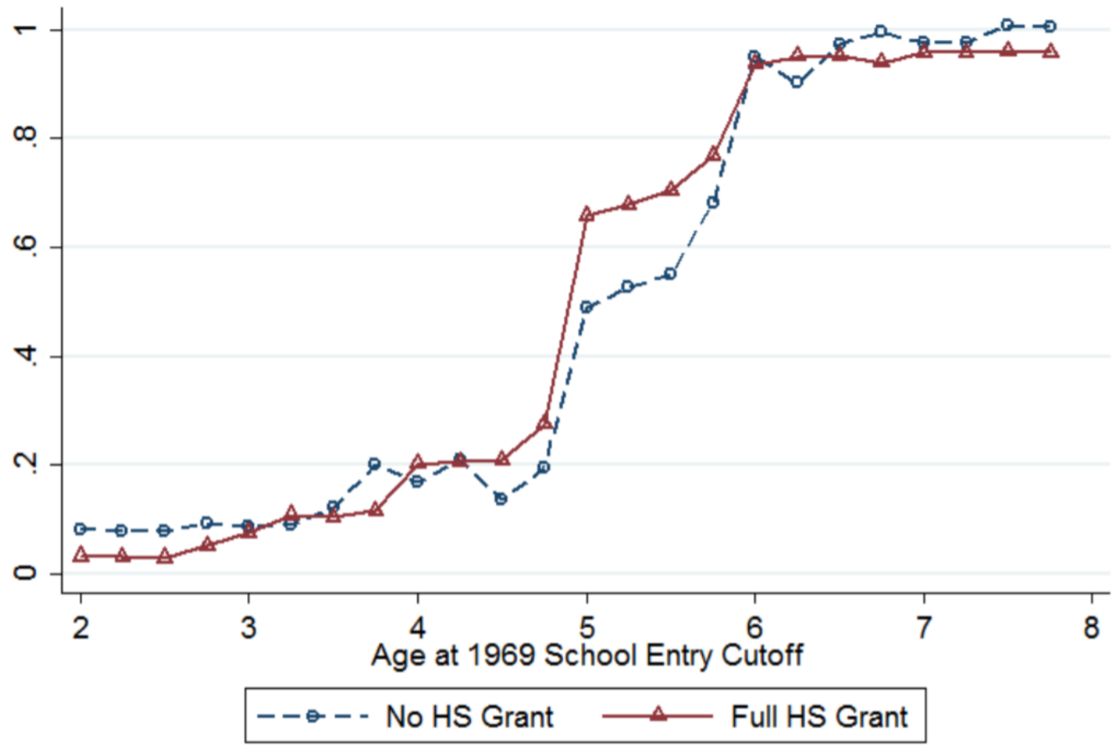
Note: Table documents the age at Head Start's launch for each birth cohort and Head Start launch date in our data. Noteworthy is that our sample is compositionally balanced from ages -1 to 15, which we present in our event study graphical analysis. Outside of those ranges, the set of counties and birth cohorts will not be compositionally balanced.

B.6 The Effect of a Head Start Launch on Head Start Enrollment

Figure A1B.1 shows the unadjusted enrollment gap in the public IPUMS data between counties that had a Head Start program in 1970 versus those that did not. Notably, four-year-old children in counties without Head Start programs were 3.4 percentage points less likely to be enrolled in school (16.8pp versus 20.2pp). Fiveyear-old children were 17 percentage points less likely to be enrolled in school (48.9 versus 65.9). In the public data, these gaps are 5.9 percentage points among 4 year olds and 21.3 percentage points among 5 year olds when looking only at children of mothers with less than a high school education.[3]

[3]Note that Head Start was not exclusively for poor kids in the 1960s and 1970s. To encourage interaction between poor children and those from less disadvantaged backgrounds, OEO policy allowed 15 percent, and later 10 percent, of children to come from families that did not meet its poverty criteria. Roughly two-thirds of children in the full-year 1969 and 1970 programs came from families in which the mother had less than a high school education, although the mothers of about 7 percent of children had attended or graduated from college.

Figure B.1: 1970 School Enrollment by Mother's Education



Notes: The figure plots the predicted school enrollment by age for children in counties with Head Start versus those in counties without Head Start in 1970. These predictions come from a linear-probability model regression using a dependent variable is equal to one if a child was enrolled in school on February 1, 1970. Because the most detailed level of geography available in the public-use data is county group, availability of Head Start is operationalized as the population-weighted share of counties in an individual's county group that had Head Start by the 1969-1970 school year. The sample is limited to states where the school-entry cutoff falls at the beginning of a quarter and includes children between school age 2 and 7 using the school-entry age cutoff date in 1969. Source: Authors' calculations using the public 1970 Census (Ruggles et al., 2017), because we have not yet disclosed this figure in the restricted 1970 Census

We explore these gaps in more detail using the restricted 1970 Census, which allows us to use a 1 in 6 sample of the U.S. population and county of residence (rather than county group). Using exact county rather than country group is more analogous to the long-term outcomes in the Census/ACS analysis. We compare school enrollment by a child's age in 1970 after adjusting for different county characteristics using either covariates or county fixed effects using the following specification:

$$SchoolEnrollment_{ic} = \boldsymbol{Z_c'}\boldsymbol{\beta_0} + \boldsymbol{A_i'}\boldsymbol{\beta_1} + \boldsymbol{A_i'}HeadStart_c\boldsymbol{\beta_2} + \epsilon_{ic} \tag{B.1}$$

where $\boldsymbol{A_i}$ has elements indicating the child's age relative to the school entry cut-off; $HeadStart_c$ is a dummy variable equal to 1 if the county had a Head Start program funded before 1970 (and is zero for places receiving their program in fiscal year 1971 or later). The set of covariates, $\boldsymbol{Z_c}$, includes either (1) $\theta_{s(c)}$, which captures state fixed effects to account for age-invariant, state-level factors that determine the local supply of preschools as well as 1960 county characteristics (share of county population in urban areas, in rural areas, under 5 years of age, 65 or older, nonwhite, with 12 or more years of education, with less than 4 years of education, in households with income less than \$3,000, in households with incomes greater than \$10,000, local government expenditures, income per capita, and whether the county was among the 300 poorest counties) or (2) county-level fixed effects (π_c). The point estimates of interest are the elements of $\boldsymbol{\beta_2}$, which, after regression-adjusting for county characteristics, capture differences in school enrollment rates of likely eligible children ages 4 to 5 in counties with Head Start.

Table B.5: Regression-Adjusted Relationship between Head Start and Enrollment

	(1)	(2)	(3)	(4)	(5)
	Dependent variable (DV): 1=School Enrollment				
	All Children			Boys	Girls
Head Start x age 3	0.038	0.038	0.038	0.037	0.039
	(0.0058)	(0.0058)	(0.0059)	(0.0065)	(0.0070)
Head Start x age 4	0.088	0.089	0.089	0.093	0.084
	(0.0088)	(0.0087)	(0.0088)	(0.0094)	(0.0095)
Head Start x age 5	0.148	0.149	0.149	0.151	0.145
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Regression-adjusted mean DV	0.52	0.56	0.52	0.52	0.52
Observations	830,000	830,000	830,000	420,000	410,000
State FE	Yes	Yes	No	No	No
County controls	No	Yes	No	No	No
County fixed effects	No	No	Yes	Yes	Yes

Notes: Sample is limited to children in states where we observe a school age entry cutoff, and where the school entry cutoff coincides with the beginning or end of a calendar quarter. Access to Head Start is measured as equal to 1 if a child lives in a county with a Head Start program in the 1969-70 school year. Observation counts are rounded per disclosure requirements. Omitted category is children age 7.75. Standard errors in brackets are clustered by county. Source: Authors calculations using the restricted 1970 Census.

The regression-adjusted, preschool enrollment gaps are summarized in Table B.5. School enrollment was 29 percent higher for all five year olds (0.149/0.52), 29 percent higher for boys (0.151/0.52) and 28 percent higher for girls (0.145/0.52). Although we are unable to assess many potential threats to the internal validity of this cross-sectional research design, the high degree of robustness of these estimates to the inclusion of different covariates in columns (2) and (3) is encouraging.

Our best estimate of the effect of a Head Start program from the 1970 Census on a birth cohort's exposure to Head Start is around 14.9 percentage points (Table B.5, column 3). It is 0.151 for men (column 4) and 0.145 for women (column 5). By construction, Census estimates should omit summer Head Start. An additional advantage of using the Census estimates is that they provide standard errors, which we use in our parametric bootstrap to scale our ITT estimates into ATETs.

Census estimates accord well with administrative data and are comparable to other studies. Administrative data suggest that in 1971, the average Head Start program served about 10 percent of resident 4-year-olds, which compares very well to around the 9 percent increase in school attendance in 1970 in the Census. The similarity of these administrative numbers and Census estimates suggests that crowd-out is minimal. To the extent that interested readers believe the estimate of the first stage should be higher or lower, they can deflate or inflate our ATET estimates accordingly.

B.7 Cost-Benefit Analysis of Head Start with the NLSY-79

A full accounting of the costs and benefits of Head Start is outside the scope of this paper. However, for comparison purposes, we compute the cumulative benefits of Head Start on economic opportunity through the program's cumulative effects on earnings potential. This is important in our context because Head Start appears to influence men and women's work effort, making the sample of wage earners selected. Our use of potential earnings follows Neal and Johnson (1996) and is directly comparable with Deming (2009). Like Deming, we use the National Longitudinal Survey of Youth 1979 (NLSY79) to predict wages for individuals born from 1957 to 1965 (ages 14 to 22 in 1979)—a time frame that overlaps our Census/ACS analysis. The NLSY data allow us to estimate the relationship between wages and components of the human capital and economic self-sufficiency indices after flexibly controlling for ability using the AFQT. Although AFQT is not available in the Census/ACS, using this as a covariate helps mitigate omitted variables bias in ability in the education and earnings relationship. We use observations on respondents' labor market wage income between 2002 and 2014, when they are between ages 35 and 57 years old, adjusted to be in 2013 dollars. We implement the following regression:

$$ln(Wages_i) = \boldsymbol{HC_i'\beta_1} + \boldsymbol{ESS_i'\beta_2} + \boldsymbol{X_i'\beta_3} + \epsilon_i \tag{B.2}$$

where $\boldsymbol{X_i}$ is a vector of race, gender, age, survey year dummy variables and a quadratic in the respondent's standardized and age-normalized AFQT score (as in Deming (2009), Neal and Johnson (1996)). In some specifications, $\boldsymbol{X_i}$ includes Deming's covariates of age 19 outcomes as regressors to account for potential sources of omitted variables bias, which has a negligible effect on our calculations. We also include new outcomes that were contained in our human capital (HC) and economic self-sufficiency indices (ESS).

Note, however, that we omit components of the ESS that are directly related to log wages such as poverty and log family income. The resulting regression coefficients capture the importance of each index component after accounting flexibly for AFQT. Table B.6 (next page) shows that the results are generally very similar with and without Deming's covariates (columns 1-3 versus columns 4-6). Panel A summarizes the internal rate of return (IRR) using the estimated cost of Head Start per student of around \$5,400 and either the (1) realized human capital and self-sufficiency gains at ages 25 to 64 or the (2) savings in public assistance outlays. We present the regression estimates underlying the calculations in (1) in panel B.

In terms of the human capital and self-sufficiency returns only, we find an IRR to Head Start of 7.7 percent averaged over men and women (Panel A, column 1). The IRR ranges from around 4 percent for women to 11 percent for men (columns 2-3), owing both to the fact that women's human capital gains are smaller and that their labor-force effort falls in response to Head Start. Note, that this approach calculates only some of the private benefits that accrue to individuals and does not include benefits through improvements in outcomes not measured here. For instance, the extent to which more education engenders better health, longevity, or well-being is ignored in these calculations.

The human capital and self-sufficiency calculations also ignore savings on public assistance expenditures. The 2000 Census and 2001-2013 ACS suggest that the average amount of dollars received by public assistance recipients between ages 25 and 54 was \$8,700 per year in 2013 dollars, which is 15 percent smaller than in the Survey of Income and Program Participation (SIPP) of \$9,967 due to misreporting. Using the SIPP calculation, the IRR on putting one child through Head Start is 2.4% overall if the only returns to Head Start were in savings in public assistance expenditures: 2.5% for men and 2.2% for women.

Adding increases in tax revenues or subtracting deadweight loss encumbered by redistributing these expenditures through the tax and welfare system would serve to increase these estimates.

Table B.6: The Effect of Human Capital and Self-Sufficiency on Adult Earnings Potential

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Ages 14 to 22			*B. Deming's sample: Ages 14 to 19*		
	Total	Male	Female	Total	Male	Female
A. Internal rate of return (IRR)						
HC and ESS only	7.7%	11.3%	4.0%	7.8%	11.6%	4.2%
Pub. assistance savings only	2.4%	2.5%	2.2%			
B. Regression estimates						
High School graduation	0.138***	0.0913*	0.217***	0.164***	0.116**	0.233***
	(0.035)	(0.048)	(0.053)	(0.041)	(0.055)	(0.065)
College completion	0.152***	0.215***	0.121***	0.158***	0.211***	0.135***
	(0.032)	(0.049)	(0.041)	(0.039)	(0.060)	(0.051)
Prof or doc degree	0.017	0.036	0.008	-0.0014	0.0228	0.003
	(0.036)	(0.060)	(0.044)	(0.043)	(0.070)	(0.054)
Professional job	0.249***	0.222***	0.256***	0.27***	0.225***	0.288***
	(0.016)	(0.025)	(0.020)	(0.020)	(0.028)	(0.024)
Years of schooling	0.030***	0.027**	0.027***	0.031***	0.0340***	0.0219*
	(0.007)	(0.011)	(0.009)	(0.009)	(0.013)	(0.012)
Weeks worked last year	0.035***	0.034***	0.036***	0.036***	0.0342***	0.0359***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Usual weekly hours	0.0199***	0.0156***	0.024***	0.020***	0.0153***	0.0240***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AFQT score	0.190***	0.183***	0.202***	0.154***	0.153***	0.156***
	(0.011)	(0.015)	(0.015)	(0.012)	(0.017)	(0.018)
AFQT squared	-0.050***	-0.036***	-0.0655***	-0.037***	-0.036***	-0.0427***
	(0.007)	(0.010)	(0.011)	(0.008)	(0.011)	(0.012)
Idle				0.044***	0.045***	0.048***
				(0.009)	(0.014)	(0.012)
Crime				0.015	0.0179*	-0.002
				(0.009)	(0.011)	(0.020)
Teen pregnancy				-0.003	-0.005	0.004
				(0.008)	(0.015)	(0.010)
Health				0.011	0.011	0.011
				(0.009)	(0.014)	(0.011)
Constant	7.843***	7.872***	6.877***	7.818***	7.772***	6.912***
	(0.0947)	(0.139)	(0.120)	(0.112	(0.161	(0.147
Observations	36,536	18,252	18,284	25,508	13,004	12,504
# Individuals	7,327	3,653	3,674	5,053	2,578	2,475
R-squared	0.409	0.356	0.400	0.414	0.366	0.405

Notes: The dependent variable is log wage income. Control variables not reported in the table include race, gender, and birth and survey year fixed effects. High school completion, college completion, and professional or doctoral degree indicate completed years of education is greater or equal to 12, 16, or 18, respectively. Heteroskedasticity-robust standard errors are beneath the point estimates in parentheses. Standard errors are clustered at the individual level to account for longitudinal dependence in the data.

B.8 Additional Estimates

The following tables present estimates by race and sex for both human capital and economic self-sufficiency. We omit these from the paper because they are imprecise for nonwhites, largely owing to the fact that nonwhite children comprise around 15 percent of the sample. In addition, smaller sample sizes for these subgroups suggest relying on the parameterized spline estimate rather than the event-study point estimate at age -1.

Table B.**7:** The Effect of Head Start on Adult Human Capital by Race

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variables	Control mean (std.)	Event study at -1 (s.e.)	ITT % Gain	Spline at -1	F-stat on break at 6 [p-value]	ATET [95% CI]	ATET %Gain
A. Whites							
Human capital index	0.01	0.0132		0.014	11.61	.089	
	(0.203)	(0.0036)			[0.00]	[.04,.154]	
High school or GED completed	0.93	0.0355	3.8	0.0021	1.479	.238	26%
	(0.076)	(0.0125)			[0.22]	[.071,.449]	
Attended some college	0.63	0.0024	0.4	0.007	4.039	.016	2.5%
	(0.144)	(0.0011)			[0.04]	[.002,.034]	
Completed 4 year college	0.3	0.0076	2.5	0.0084	7.918	.051	17%
	(0.127)	(0.0023)			[0.00]	[.02,.091]	
Prof. or doc. Degree	0.029	0.0063	21.7	0.00175	6.957	.042	145%
	(0.039)	(0.0026)			[0.01]	[.008,.085]	
Years of schooling	13.64	0.002	0.0	0.0441	10.46	.013	0.10%
	(0.708)	(0.0006)			[0.00]	[.005,.024]	
Has a professional job	0.36	0.0067	1.9	0.00658	4.645	.045	12.5%
	(0.126)	(0.0024)			[0.03]	[.013,.085]	
B. Nonwhites							
Human capital index	0.015	0.0037		0.00147	5.278	.025	
	(0.33)	(0.0075)			[0.02]	[-.076,.13]	
High school or GED completed	0.86	-0.003	-0.3	-0.00004	2.134	-.02	-2.3%
	(0.16)	(0.0257)			[0.14]	[-.37,.34]	
Attended some college	0.56	-0.00018	0.0	0.000693	1.759	-.001	-0.2%
	(0.23)	(0.0035)			[0.18]	[-.049,.047]	
Completed 4 year college	0.19	0.00056	0.3	0.00126	5.126	.004	2.1%
	(0.18)	(0.0049)			[0.02]	[-.063,.073]	
Prof. or doc. Degree	0.017	0.004	23.5	0.00077	1.9	.027	159%
	(0.06)	(0.0044)			[0.17]	[-.031,.091]	
Years of schooling	13	0.0008	0.0	-0.007	2.712	.005	0.038%
	(1.16)	(0.0014)			[0.10]	[-.013,.025]	
Has a professional job	0.26	0.0028	1.1	0.000413	1.629	.019	7.3%
	(0.20)	(0.0049)			[0.20]	[-.047,.089]	

Notes: In column 1, the control mean and standard deviation are calculated using the cohorts ages 6 and 7 at the time Head Start was launched. Column 2 presents the estimated intention-to-treat (ITT) effect evaluated at birth cohort of full exposure (-1, see Figure 2). Columns 3 and 7 compute the percentage increase implied by the ITT or ATET, respectively, estimate relative to the control mean (the ratio of column 2 or 6 to column 1) for components of the index. Column 4 presents the ITT spline estimate evaluated at -1. Column 5 presents the F-statistic and p-value for the test of a trend-break in the spline at age 6. The ATET estimate in column 6 divides the ITT effect at -1 by the Bailey, Sun, and Timpe Online Appendix – 14 estimate of receiving a Head Start grant on school enrollment at school age 5—0.149 (s.e. 0.022) for the full sample, see Appendix Table B.5).

Table B.8: The Effect of Head Start on Adult Human Capital by Race: White Men and Women

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables	Control mean (std.)	Event study at -1 (s.e.)	ITT % Gain	Spline at -1	F-stat on break at 6 [p-value]	ATET [95% CI]
A. White females						
Human capital index	0.011	0.0066		0.0098	3.57	.046
	(0.2328)	(0.0044)			[0.06]	[-.014,.115]
High school or GED completed	0.94	0.0017	0.2	0.00175	0.4132	.012
	(0.0827)	(0.0013)			[0.52]	[-.006,.032]
Attended some college	0.66	0.0038	0.6	0.00476	0.2301	.026
	(0.1668)	(0.0027)			[0.63]	[-.01,.069]
Completed 4 year college	0.31	0.0017	0.5	0.00546	1.8338	.012
	(0.1538)	(0.0032)			[0.18]	[-.032,.059]
Prof. or doc. Degree	0.023	0.0011	4.8	0.00084	3.6028	.008
	(0.0441)	(0.0007)			[0.06]	[-.002,.019]
Years of schooling	13.69	0.0114	0.1	0.0301	2.9967	.079
	(0.8099)	(0.0151)			[0.08]	[-.128,.306]
Has a professional job	0.38	0.0041	1.1	0.00483	3.1632	.028
	(0.1559)	(0.0031)			[0.08]	[-.014,.077]
B. White males						
Human capital index	0.0098	0.0198		0.0182	11.78	.131
	(0.2515)	(0.0043)			[0.00]	[.072,.213]
High school or GED completed	0.92	0.0033	0.4	0.00259	1.7827	.022
	(0.1016)	(0.0016)			[0.18]	[.001,.047]
Attended some college	0.6	0.0115	1.9	0.0098	6.1983	.076
	(0.1791)	(0.0030)			[0.01]	[.036,.13]
Completed 4 year college	0.3	0.0108	3.6	0.0112	9.0044	.072
	(0.1560)	(0.0029)			[0.00]	[.033,.124]
Prof. or doc. Degree	0.036	0.0029	8.1	0.0028	3.9562	.019
	(0.0557)	(0.0009)			[0.05]	[.007,.035]
Years of schooling	13.59	0.0594	0.4	0.0588	11.0703	.393
	(0.8919)	(0.0146)			[0.00]	[.196,.663]
Has a professional job	0.35	0.0095	2.7	0.0091	2.8856	.063
	(0.1584)	(0.0027)			[0.09]	[.027,.11]

Notes: See Table B.7 notes.

Table B.9: The Effect of Head Start on Adult Human Capital by Race: Nonwhite Men and Women

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables	Control mean (std.)	Event study at -1 (s.e.)	ITT % Gain	Spline at -1	F-stat on break at 6 [p-value]	ATET [95% CI]
C. Nonwhite females						
Human capital index	0.015	0.009		0.00364	7.11	.062
	(0.3668)	(0.0091)			[0.01]	[-.062,.201]
High school or GED completed	0.87	-0.000041	0.0	0.00091	3.6186	0
	(0.1783)	(0.0042)			[0.06]	[-.06,.06]
Attended some college	0.6	0.0052	0.9	0.00301	3.4384	.036
	(0.2627)	(0.0061)			[0.06]	[-.048,.128]
Completed 4 year college	0.2	0.0088	4.4	0.00434	6.7319	.061
	(0.2117)	(0.0054)			[0.01]	[-.012,.147]
Prof. or doc. Degree	0.016	-0.00031	-1.9	0.0000098	0.5994	-.002
	(0.0640)	(0.0017)			[0.44]	[-.026,.022]
Years of schooling	13.1	0.0198	0.2	-0.00329	3.1776	.137
	(1.2751)	(0.0306)			[0.07]	[-.284,.594]
Has a professional job	0.29	0.0071	2.4	0.00147	3.381	.049
	(0.2335)	(0.0069)			[0.07]	[-.045,.155]
C. Nonwhite males						
Human capital index	0.016	-0.0039		-0.000322	0.546	-.026
	(0.4032)	(0.0105)			[0.46]	[-.171,.116]
High school or GED completed	0.85	-0.0014	-0.2	-0.00077	0.1913	-.009
	(0.2037)	(0.0051)			[0.66]	[-.079,.06]
Attended some college	0.51	-0.0054	-1.1	-0.00189	0.0057	-.036
	(0.2839)	(0.0072)			[0.94]	[-.138,.061]
Completed 4 year college	0.18	-0.0014	-0.8	-0.00182	0.3928	-.009
	(0.2188)	(0.0059)			[0.53]	[-.09,.071]
Prof. or doc. Degree	0.018	0.0014	7.8	0.00147	1.1057	.009
	(0.0739)	(0.0021)			[0.29]	[-.018,.039]
Years of schooling	12.87	-0.0352	-0.3	-0.0077	0.9593	-.233
	(1.4293)	(0.0356)			[0.33]	[-.746,.241]
Has a professional job	0.22	-0.00059	-0.3	0.000294	0.0229	-.004
	(0.2324)	(0.0062)			[0.88]	[-.088,.081]

Notes: See Table B.7 notes.

Table B.10: The Effect of Head Start on Adult Self-Sufficiency by Race

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables	Control mean (std.)	Event study at -1 (s.e.)	ITT % Gain	Spline at -1	F-stat on break at 6 [p-value]	ATET [95% CI]
A. Whites						
Economic self-sufficiency index	0.021	0.0055		0.0041	5.328	.037
	(0.2)	(0.0024)			[0.02]	[.005,.076]
Worked last year	0.87	0.001	0.11	0.0010	0.4083	.007
	(0.1)	(0.0014)			[0.52]	[-.012,.027]
No. weeks worked last year	41.6	0.0558	0.13	0.0539	0.6303	.374
	(5.2)	(0.0758)			[0.43]	[-.636,1.47]
Usual hours works per week	36.2	0.0229	0.06	0.0455	1.275	.154
	(4.9)	(0.0804)			[0.26]	[-1.08,1.49]
In poverty*	0.08	-0.0015	-1.9	-0.0016	5.059	-.01
	(0.1)	(0.0011)			[0.02]	[-.027,.005]
Rec'd income from public sources*	0.1	-0.0052	-5.2	-0.0040	13.22	-.035
	(0.1)	(0.0010)			[0.00]	[-.056,-.02]
B. Nonwhites						
Economic self-sufficiency index	0.018	-0.0017		0.0070	1.188	-.011
	(0.28)	(0.0056)			[0.28]	[-.089,.066]
Worked last year	0.79	-0.0022	-0.28	0.0021	0.1255	-.015
	(0.20)	(0.0038)			[0.72]	[-.069,.037]
No. weeks worked last year	37.0	-0.069	-0.19	0.1792	1.211	-.463
	(10)	(0.2000)			[0.27]	[-3.255,2.29]
Usual hours works per week	32.4	-0.0551	-0.17	0.1099	0.2172	-.37
	(9.38)	(0.1820)			[0.64]	[-2.894,2.145]
Not in poverty	0.2	-0.0004	-0.20	-0.0020	0.9526	-.003
	(0.19)	(0.0038)			[0.33]	[-.055,.05]
No income from public sources	0.19	0.0013	0.68	-0.0003	0.0164	.009
	(0.19)	(0.0035)			[0.90]	[-.038,.058]

Notes: See Table B.7 notes.

Table B.11: The Effect of Head Start on Adult Self-Sufficiency by Race and Race-Sex: White Women and Men

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables	Control mean (std.)	Event study at -1 (s.e.)	ITT % Gain	Spline at -1	F-stat on break at 6 [p-value]	ATET [95% CI]
A. White females						
Economic self-sufficiency index	0.015	0.0063		0.0041	2.95	.043
	(0.1834)	(0.0030)			[0.09]	[.003,.093]
Worked last year	0.79	-0.0021	-0.3	-0.0022	0.2406	-.014
	(0.1346)	(0.0023)			[0.62]	[-.049,.018]
No. weeks worked last year	37.96	-0.123	-0.3	-0.1120	0.2377	-.848
	(6.9050)	(0.1160)			[0.63]	[-2.628,.761]
Usual hours works per week	30.56	-0.148	-0.5	-0.1197	0.1334	-1.021
	(5.9220)	(0.1040)			[0.72]	[-2.665,.414]
Log labor income wage only	10.31	0.0039		0.0048	2.33	.027
	(0.3501)	(0.0056)			[0.13]	[-.05,.111]
Log family income/poverty line	5.87	0.0069		0.0059	1.5198	.048
	(0.3097)	(0.0058)			[0.22]	[-.031,.137]
Not in poverty	0.9	0.0022	0.2	0.0025	5.966	.015
	(0.1026)	(0.0017)			[0.01]	[-.008,.042]
No income from public sources	0.89	0.0047	0.5	0.0025	0.7188	.032
	(0.1128)	(0.0015)			[0.40]	[.012,.059]
B. White males						
Economic self-sufficiency index	0.029	0.0045		0.0042	1.81	.030
	(0.2007)	(0.0031)			[0.18]	[-.01,.076]
Worked last year	0.92	0.004	0.4	0.0034	4.2597	.026
	(0.1011)	(0.0013)			[0.04]	[.009,.048]
No. weeks worked last year	45.5	0.195	0.4	0.1750	4.7986	1.291
	(5.7792)	(0.0694)			[0.03]	[.379,2.443]
Usual hours works per week	42.01	0.156	0.4	0.1694	5.54	1.033
	(5.8846)	(0.0818)			[0.02]	[-.025,2.292]
Log labor income wage only	10.92	0.0011		0.0014	0.0608	.007
	(0.3009)	(0.0046)			[0.81]	[-.054,.071]
Log family income/poverty line	5.95	0.00071		-0.0005	0.0218	.005
	(0.2961)	(0.0052)			[0.88]	[-.065,.077]
Not in poverty	0.94	0.00049	0.1	0.0002	0.0754	.003
	(0.0884)	(0.0013)			[0.78]	[-.014,.021]
No income from public sources	0.9	0.0053	0.6	0.0053	18.34	.035
	(0.1081)	(0.0014)			[0.00]	[.016,.06]

Notes: See Table B.7 notes.

Table B.12: The Effect of Head Start on Adult Self-Sufficiency by Race and Race-Sex: Nonwhite Women and Men

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables	Control mean (std.)	Event study at -1 (s.e.)	ITT % Gain	Spline at -1	F-stat on break at 6 [p-value]	ATET [95% CI]
C. Nonwhite females						
Economic self-sufficiency index	0.017	0.0016		0.0067	1.45	.011
	(0.3127)	(0.0070)			[0.23]	[-.087,.113]
Worked last year	0.75	-0.0019	-0.3	0.0032	0.8274	-.013
	(0.2278)	(0.0052)		)	[0.36]	[-.088,.061]
No. weeks worked last year	35.56	0.003	0.0	0.1855	1.5066	.021
	(11.8287)	(0.2690)			[0.22]	[-3.772,3.9]
Usual hours works per week	30.05	-0.0199	-0.1	0.0959	0.4792	-.137
	(10.1554)	(0.2280)			[0.49]	[-3.37,3.122]
Log labor income wage only	10.25	0.0014		0.0053	1.3066	.01
	(0.5443)	(0.0129)			[0.25]	[-.171,.196]
Log family income/poverty line	5.39	-0.0099		-0.0015	0.1804	-.068
	(0.5615)	(0.0141)			[0.67]	[-.278,.129]
Not in poverty	0.77	0.0016	0.2	0.0012	1.5023	.011
	(0.2220)	(0.0052)			[0.22]	[-.061,.087]
No income from public sources	0.8	-0.00008	0.0	0.0028	0.3556	-.001
	(0.2184)	(0.0044)			[0.55]	[-.063,.063]
D. Nonwhite males						
Economic self-sufficiency index	0.017	-0.0113		0.0049	0.055	-.075
	(0.3441)	(0.0083)			[0.82]	[-.198,.035]
Worked last year	0.82	-0.0066	-0.8	0.0012	0.0741	-.044
	(0.2209)	(0.0056)			[0.79]	[-.126,.031]
No. weeks worked last year	38.84	-0.334	-0.9	0.0490	0.0002	-2.212
	(12.1310)	(0.2920)			[0.99]	[-6.488,1.661]
Usual hours works per week	35.43	-0.274	-0.8	0.0105	0.1362	-1.815
	(11.4734)	(0.2750)			[0.71]	[-5.778,1.849]
Log labor income wage only	10.56	-0.00021		-0.0042	0.2214	-.001
	(0.5432)	(0.0134)			[0.64]	[-.182,.183]
Log family income/poverty line	5.57	0.0019		0.0036	0.1884	.013
	(0.5623)	(0.0124)			[0.66]	[-.154,.185]
Not in poverty	0.83	-0.003	-0.4	0.0016	0.0131	-.02
	(0.2156)	(0.0051)			[0.91]	[-.091,.049]
No income from public sources	0.82	-0.0026	-0.3	-0.0028	0.1825	-.017
	(0.2218)	(0.0053)			[0.67]	[-.091,.054]

Notes: See Table B.7 notes.

Table B.13: The Effect of Head Start on Incarceration and Mortality

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables	Control mean (std.)	Event study at -1 (s.e.)	ITT % Gain	Spline at -1	F-stat on break at 6 [p-value]	ATET [95% CI]
A. Full Sample						
Survived to 2000	0.98	-0.000089	-1.6	-0.0001	0.404	-.001
	(0.0057)	(0.0003)			[0.52]	[-.004,.003]
Incarcerated	0.01	0.00017	1.7	0.0002	0.111	.001
	(0.0277)	(0.0005)			[0.74]	[-.005,.008]
C. White females						
Survived to 2000	0.99	0.000016	0.4	0.0000	0.178	0
	(0.0044)	(0.0002)			[0.67]	[-.002,.003]
Incarcerated	0.00	0.00028	n/a	0.0003	2.62	.002
	(0.0140)	(0.0003)			[0.11]	[-.002,.006]
C. White males						
Survived to 2000	0.97	0.00022	2.7	0.0003	0.767	.001
	(0.0083)	(0.0003)			[0.38]	[-.003,.006]
Incarcerated	0.01	0.00019	1.9	0.0001	2.62	.001
	(0.0409)	(0.0008)			[0.59]	[-.009,.012]
D. Nonwhite females						
Survived to 2000	0.98	0.0003	2.7	-0.0001	3.43	.002
	(0.0111)	(0.0005)			[0.06]	[-.004,.009]
Incarcerated	0.01	0.00112	11.2	-0.0002	2.62	.008
	(0.0486)	(0.0017)			[0.62]	[-.016,.033]
E. Nonwhite males						
Survived to 2000	0.96	0.0000099	0.1	-0.0006	0.862	0
	(0.0176)	(0.0008)			[0.35]	[-.011,.011]
Incarcerated	0.08	0.000055	0.1	0.0014	2.62	0
	(0.1758)	(0.0062)			[0.88]	[-.083,.086]

Notes: See Table B.7 notes.

Table B.14: The Effect of Head Start on the Human Capital Index using Different Measures of Access

	(1) This paper (spline at -1)	(2) Barr and Gibbs (2017) differences-in-differences	(3) Years of access	(4) Thompson (2017) funding per capita	(5) Johnson and Jackson (2017) funding per capita
A. Excluding State-by-Birth-Year Fixed Effects					
Effect of Head Start	0.0152	0.0152	0.0191	0.00084	0.00130
	(0.00064)	(0.00477)	(0.00571)	(0.000458)	(0.000562)
B. Including State-by-Birth-Year Fixed Effect					
Effect of Head Start	0.0154	0.0110	0.0147	0.000347	0.000724
	(0.003)	(0.00256)	(0.00317)	(0.000439)	(0.000395)

Notes: There are 14,800,000 individuals (rounded for disclosure) in each regression. Panel A presents the results from specifications that exclude state-by-year-ofbirth fixed effects as is necessary in longitudinal samples. Panel B presents the results from specifications that exclude state-by-year-of-birth fixed effects as is necessary in longitudinal samples. The only change in the specification is the specification of the variable used to measure access to Head Start. Column 1 repeats the value shown in the paper. Column 2 presents a differences-in-differences specification, where Head Start=1 in a county for all children younger than 6 when it began. Column 3 uses a specification that measures share of the three potential years of access in which a child lived in a county with a Head Start program and was age eligible (possible variable values are 0, 1/3, 2/3, and 1). Column 4 uses the Thompson (2017) measure of Head Start access as average per-year, per-capita Head Start spending in the three years when an individual was between 3 and 6 years old. Column 5 uses the Johnson and Jackson (2017) measure of Head Start access as Head Start spending per poor 4-year-old, as defined when the cohort was 4 years old.

Table B.15: The Effect of Head Start on the Self-Sufficiency Index using Different Measures of Access

	(1) This paper (spline at -1)	(2) Barr and Gibbs (2017) differences-in-differences	(3) Years of access	(4) Thompson (2017) funding per capita	(5) Johnson and Jackson (2017) funding per capita
A. Excluding State-by-Birth-Year Fixed Effects					
Effect of Head Start	0.0055	0.00397	0.00488	0.000126	0.000193
	(0.00026)	(0.00146)	(0.00159)	(0.000201)	(0.000156)
B. Including State-by-Birth-Year Fixed Effect					
Effect of Head Start	0.0048	0.00232	0.00296	-0.000318	-1.85e-05
	(0.00022)	(0.00124)	(0.00147)	(0.000241)	(0.000119)

Notes: See Table B.14 notes.

Table B.16: The Effect of Head Start on Completed High School or GED using Different Measures of Access

	(1) This paper (spline at -1)	(2) Barr and Gibbs (2017) differences-in-differences	(3) Years of access	(4) Thompson (2017) funding per capita	(5) Johnson and Jackson (2017) funding per capita
A. Excluding State-by-Birth-Year Fixed Effects					
Effect of Head Start	0.0028	0.00357	0.00451	0.000164	0.000231
	(0.00026)	(0.00120)	(0.00140)	(0.000148)	(0.000108)
B. Including State-by-Birth-Year Fixed Effect					
Effect of Head Start	0.00245	0.00227	0.00305	-8.47e-05	6.73e-05
	(0.001)	(0.000673)	(0.000834)	(0.000144)	(7.52e-05)

Notes: See Table B.14 notes.

Table B.17: The Effect of Head Start on Enrolled in College using Different Measures of Accesss

	(1) This paper (spline at -1)	(2) Barr and Gibbs (2017) differences-in-differences	(3) Years of access	(4) Thompson (2017) funding per capita	(5) Johnson and Jackson (2017) funding per capita
A. Excluding State-by-Birth-Year Fixed Effects					
Effect of Head Start	0.0077	0.00949	0.0117	0.000309	0.000652
	(0.0006)	(0.00384)	(0.00467)	(0.000324)	(0.000299)
B. Including State-by-Birth-Year Fixed Effect					
Effect of Head Start	0.0077	0.00564	0.00733	5.23e-05	0.000211
	(0.002)	(0.00136)	(0.00175)	(0.000302)	(0.000170)

Notes: See Table B.14 notes.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge. 2017. "When should you adjust standard errors for clustering?" NBER Working Paper 24003.

Agostinelli, Francesco, and Giuseppe Sorrenti. 2018. "Money vs. time: Family income, maternal labor supply, and child development." HCEO Working Paper 2018-017.

Ahammer, Alexander, Martin Halla, and Nicole Schneeweis. 2018. "The Effect of Prenatal Maternity Leave on Short and Long-Term Child Outcomes." IZA DP No. 11394.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. "The long-run impact of cash transfers to poor families." *The American Economic Review*, 106(4): 935–971.

Almond, Douglas, and Janet Currie. 2011. "Human Capital Development before Age Five." Elsevier Handbook of Labor Economics.

Almond, Douglas, Hilary W. Hoynes, and Diane Whitmore Schanzenbach. 2011. "Inside the War on Poverty: The impact of Food Stamps on birth outcomes." *The Review of Economics and Statistics*, 93(2): 387–403.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature*, 56(4): 1360–1446.

American Academy of Pediatrics, and Pediatric Policy Council. 2015. "Major pediatric associations call for congressional action on paid leave."

Applebaum, Eileen, and Ruth Milkman. 2011. "Leaves that pay: Employer and work experiences with paid family leave in California." Center for Economic and Policy Research.

Arellano, M. 1987. "Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics*, 49(4): 431–434.

Asai, Yukiko. 2015. "Parental leave reforms and the employment of new mothers: Quasi-experimental evidence from Japan." *Labour Economics*, 36: 72–83.

Asai, Yukiko. 2019. "Costs of Employment and Flexible Labor Demand: Evidence from Maternity and Parental Leave Reforms." Social Science Research Network SSRN Scholarly Paper ID 3362488, Rochester, NY.

Autor, David, Mark Duggan, Jonathan Gruber, and Catherine Maclean. 2013. "How does Access to Short Term Disability Insurance Impact SSDI Claiming?" National Bureau of Economic Research.

Averett, Susan L., and Leslie A. Whittington. 2001. "Does Maternity Leave Induce Births?" *Southern Economic Journal*, 68(2): 403–417.

Bailey, Martha J. 2006. "More power to the pill: The impact of contraceptive freedom on women's life cycle labor supply." *Quarterly Journal of Economics*, 121(1): 289–320.

Bailey, Martha J. 2012. "Reexamining the Impact of Family Planning Programs on US Fertility: Evidence from the War on Poverty and the Early Years of Title X." *American Economic Journal: Applied Economics*, 4(2): 62–97.

Bailey, Martha J., and Andrew Goodman-Bacon. 2015. "The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans." *American Economic Review*, 105(3): 1067–1104.

Bailey, Martha J., and Nicolas J. Duquette. 2014. "How Johnson Fought the War on Poverty: The Economics and Politics of Funding at the Office of Economic Opportunity." *The journal of economic history*, 74(2): 351–388.

Bailey, Martha J., and Sheldon Danziger. 2013. "The Legacies of the War on Poverty." In *Legacies of the War on Poverty*. New York:Russell Sage Foundation.

Bailey, Martha J., Brad Hershbein, and Amalia R. Miller. 2012. "The opt-in revolution? Contraception and the gender gap in wages." *American Economic Journal: Applied Economics*, 4(3): 225–54.

Bailey, Martha J., Hilary Hoynes, Maya Rossin-Slater, and Reed Walker. 2019. "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program."

Bailey, Martha J., Shuqiao Sun, and Brenden Timpe. 2018. "Prep school for poor kids: The long-run impacts of Head Start on human capital and economic self-sufficiency."

Baker, Michael, and Kevin Milligan. 2008. "How does job-protected maternity leave affect mothers' employment?" *Journal of Labor Economics*, 26(4): 655–691.

Baker, Michael, and Kevin Milligan. 2010. "Evidence from Maternity Leave Expansions of the Impact of Maternal Care on Early Child Development." *Journal of Human Resources*, 45(1): 1–32.

Baker, Michael, and Kevin Milligan. 2014. "Maternity leave and children's cognitive and behavioral development." *Journal of Population Economics*, 28(2): 373–391.

Bana, Sarah, Kelly Bedard, and Maya Rossin-Slater. 2018. "The impacts of paid family leave benefits: regression kink evidence from California administrative data." IZA DP No. 11381.

Barnett, W.S., and Leonard N. Masse. 2007. "Comparative benefit–cost analysis of the Abecedarian program and its policy implications." *Economics of Education Review*, 26(1): 113–125.

Barr, Andrew, and Chloe R Gibbs. 2017. "Breaking the Cycle? Intergenerational Effects of an Anti-Poverty Program in Early Childhood."

Bartel, Ann P., Maya Rossin-Slater, Christopher J. Ruhm, Jenna Stearns, and Jane Waldfogel. 2018. "Paid Family Leave, Fathers' Leave-Taking, and Leave-Sharing in Dual-Earner Households." *Journal of Policy Analysis and Management*, 37(1): 10–37.

Bassok, Daphna, Maria Fitzpatrick, and Susanne Loeb. n.d.. "Does state preschool crowd-out private provision? The impact of universal pre-kindergarten on the childcare sector in Oklahoma and Georgia." *Journal of Urban Economics*, 83: 18–33.

Bastian, Jacob. 2018. "The rise of working mothers and the 1975 Earned Income Tax Credit."

Bastian, Jacob, and Kathy Michelmore. 2018. "The long-term impact of the Earned Income Tax Credit on children's education and employment outcomes." *Journal of Labor Economics*, 36(4): 1127–1163.

Baum, Charles L. 2003. "The effect of state maternity leave legislation and the 1993 Family and Medical Leave Act on employment and wages." *Labour Economics*, 10(5): 573–596.

Baum, Charles L., and Christopher J. Ruhm. 2016. "The effects of paid family leave in California on labor market outcomes." *Journal of Policy Analysis and Management*, 35(2): 333–356.

Beaudry, Paul, and Ethan Lewis. 2014. "Do Male-Female Wage Differentials Reflect Differences in the Return to Skill? Cross-City Evidence from 1980-2000." *American Economic Journal: Applied Economics*, 6(2): 178–194.

Becker, Gary S. 1975. "Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition."

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, 75(4, Part 1): 352–365.

Berger, Lawrence M., and Jane Waldfogel. 2004. "Maternity leave and the employment of new mothers in the United States." *Journal of Population Economics*, 17(2): 331–349.

Berman, Eli, John Bound, and Zvi Griliches. 1994. "Changes in the Demand for Skilled Labor within U. S. Manufacturing: Evidence from the Annual Survey of Manufactures." *The Quarterly Journal of Economics*, 109(2): 367–397.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics*, 2(3): 228–255.

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

Black, Dan A., Amelia M. Haviland, Seth G. Sanders, and Lowell J. Taylor. 2008. "Gender Wage Disparities among the Highly Educated." *Journal of Human Resources*, 43(3): 630–659.

Blatter, Marc, Samuel Muehlemann, and Samuel Schenker. 2012. "The costs of hiring skilled workers." *European Economic Review*, 56: 20–35.

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. 1997. "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s." *Journal of Labor Economics*, 15(1): 1–42.

Blau, Francine D., and Lawrence M. Kahn. 2006. "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence." *Industrial and Labor Relations Review*, 60(1): 45–66.

Blau, Francine D., and Lawrence M. Kahn. 2013. "Female Labor Supply: Why Is the United States Falling Behind?" *American Economic Review*, 103(3): 251–56.

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

Bloom, Benjamin S. 1964. *Stability and change in human characteristics.* New York:Wiley. OCLC: 224354.

Bobo, Lawrence, James Johnson, Barry Bluestone, Irene Browne, Sheldon Danziger, Philip Moss, Gary P. Green, Harry Holzer, Joleen Kirschenman, Maria Krysan, Camille Zubrinsky Charles, Michael Massagli, Melvin Oliver, Reynolds Farley, and Chris Tilly. 2008. "Multi-City Study of Urban Inequality, 1992-1994: [Atlanta, Boston, Detroit, and Los Angeles]."

Bound, John, and Gary Solon. 1999. "Double trouble: on the value of twins-based estimation of the return to schooling." *Economics of Education Review*, 18(2): 169–182.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement error in survey data." *Handbook of Labor Economics*, 5: 3705–3843.

Braun, Samuel J, and Esther P Edwards. 1972. *History and theory of early childhood education.* Worthington, Ohio, C.A. Jones Pub. Co. OCLC: 639993516.

Brenøe, Anne A., Serena Canaan, Nikolaj A. Harmon, and Heather Royer. 2018. "Is parental leave costly for firms and coworkers?"

Bronson, Mary Ann. 2015. "Degrees Are Forever: Marriage, Educational Investment, and Lifecycle Labor Decisions of Men and Women."

Brown, David W., Amanda E. Kowalski, and Ithai Z. Lurie. 2015. "Medicaid as an investment in children: What is the long-term impact on tax receipts?" NBER Working Paper No. 20835.

Buckles, Kasey S., and Daniel M. Hungerman. 2013. "Season of birth and later outcomes: Old questions, new answers." *The Review of Economics and Statistics*, 95(3): 711–724.

Byker, Tanya S. 2016. "Paid Parental Leave Laws in the United States: Does Short-Duration Leave Affect Women's Labor-Force Attachment?" *American Economic Review*, 106(5): 242–46.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A practitioner's guide to cluster-robust inference." *Journal of Human Resources*, 50(2): 317–372.

Card, David. 1992. "Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage." *Industrial and Labor Relations Review*, 46(1): 22–37.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *The Quarterly Journal of Economics*, 131(2): 633–686.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica*, 83(6): 2453–2483.

Carneiro, Pedro, and Rita Ginja. 2014. "Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start." *American Economic Journal: Economic Policy*, 6(4): 135–173.

Carneiro, Pedro, Katrine V. Løken, and Kjell G. Salvanes. 2015. "A flying start? Maternity leave benefits and long-run outcomes of children." *Journal of Political Economy*, 123(2): 365–412.

Cascio, Elizabeth U., and Diane Whitmore Schanzenbach. 2013. "The Impacts of Expanding Access to High-Quality Preschool Education." *Brookings Papers on Economic Activity*, 127–178.

Chung, YoonKyung, Barbara Downs, Danielle H. Sandler, and Robert Sienkiewicz. 2017. "The Parental Gender Earnings Gap in the United States." Census Bureau Center for Economic Studies Working Paper 17-68.

Correll, Shelley J., Stephen Benard, and In Paik. 2007. "Getting a job: Is there a motherhood penalty?" *American Journal of Sociology*, 112(5): 1297–1338.

Cortés, Patricia, and Jessica Pan. 2018. "When Time Binds: Substitutes for Household Production, Returns to Working Long Hours, and the Skilled Gender Wage Gap." *Journal of Labor Economics*, 37(2): 351–398.

Cunha, Flavio, and James Heckman. 2007. "The technology of skill formation." *American Economic Review*, 97(2): 31–47.

Currie, Janet. 2001. "Early Childhood Education Programs." *Journal of Economic Perspectives*, 15(2): 213–238.

Currie, Janet, and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review*, 85(3): 341–364.

Dahl, Gordon B., and Lance Lochner. 2012. "The impact of family income on child achievement: Evidence from the Earned Income Tax Credit." *American Economic Review*, 102(5): 1927–56.

Dahl, Gordon B., Katrine V. Løken, Magne Mogstad, and Kari Vea Salvanes. 2016. "What is the case for paid maternity leave?" *Review of Economics and Statistics*, 98(4): 655–670.

Danzer, Natalia, and Victor Lavy. 2018. "Paid parental leave and children's schooling outcomes." *The Economic Journal*, 128(608): 81–117.

Das, Tirthatanmoy, and Solomon W. Polachek. 2015. "Unanticipated effects of California's paid family leave program." *Contemporary Economic Policy*, 33(4): 619–635.

DeLauro, Rosa, and Kirsten Gillibrand. 2019. "DeLauro, Gillibrand Reintroduce the FAMILY Act."

Deming, David. 2009. "Early childhood intervention and life-cycle skill development: Evidence from Head Start." *American Economic Journal: Applied Economics*, 1(3): 111–134.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. "Chapter 61 Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*. Vol. 4, , ed. T. Paul Schultz and John A. Strauss, 3895–3962. Elsevier.

Duncan, Greg J., and Katherine Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives*, 27(2): 109–132.

Dustmann, Christian, and Uta Schönberg. 2012. "Expansions in Maternity Leave Coverage and Children's Long-Term Outcomes." *American Economic Journal: Applied Economics*, 4(3): 190–224.

Efron, Bradley, and Robert J Tibshirani. 1993. *An introduction to the bootstrap.* New York, N.Y.; London:Chapman & Hall. OCLC: 797437299.

Espinosa, Linda M. 2002. "High-Quality Preschool: Why We Need It and What It Looks Like." 14.

Farré, Lidia, and Libertad Gonzalez. 2019. "Does Paternity Leave Reduce Fertility?" Social Science Research Network SSRN Scholarly Paper ID 3318797, Rochester, NY.

Faulkner, Edwin J. 1940. *Accident-and-Health Insurance.* New York and London:McGraw-Hill Book Company Inc.

Fosburg, Linda B, Nancy N Goodrich, Mary Kay Fox, Patricia Granahan, Janet Smith, John H Rimes, and Michael Weitzman. n.d.. "The Effects of Head Start Health Services: Report of the Head Start Health Evaluation." In *Report Prepared for the ADministration for Children, Youth and Families, U.S. Department of Health and Human Services.* 1191. Cambridge, MA:Abt Associates.

Frisvold, David E. 2015. "Nutrition and cognitive achievement: An evaluation of the School Breakfast Program." *Journal of Public Economics*, 124: 91–104.

Fuchs, Victor R. 1989. "Women's Quest for Economic Equality." *The Journal of Economic Perspectives*, 3(1): 25–41.

Gallen, Yana. 2018. "The effect of maternity leave extensions on firms and coworkers." Working paper.

Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. "Longer-Term Effects of Head Start." *American Economic Review*, 92(4): 999–1012.

Gladstone, Leslie W., Jennifer D. Williams, and Richard S. Belous. 1985. "Maternity and parental leave policies: A comparative analysis." Congressional Research Service 85-184 GOV.

Goldin, Claudia. 1988. "Marriage Bars: Discrimination Against Married Women Workers, 1920's to 1950's." National Bureau of Economic Research Working Paper 2747.

Goldin, Claudia. 1994. "Labor Markets in the Twentieth Century." National Bureau of Economic Research Working Paper 58.

Goldin, Claudia. 2014. "A grand gender convergence: Its last chapter." *The American Economic Review*, 104(4): 1091–1119.

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

Goldin, Claudia, and Lawrence F. Katz. 2002. "The power of the pill: Oral contraceptives and women's career and marriage decisions." *Journal of Political Economy*, 110(4): 730–770.

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." National Bureau of Economic Research Working Paper 12139.

Goodman-Bacon, Andrew. 2018. "Public Insurance and Mortality: Evidence from Medicaid Implementation." *Journal of Political Economy*, 126(1): 216–262.

Griliches, Zvi. 1979. "Sibling Models and Data in Economics: Beginnings of a Survey." *Journal of Political Economy*, 87(5, Part 2): S37–S64.

Grosz, Michel Z., Douglas L. Miller, and Na'ama Shenhav. 2017. "All In the Family: Assessing the External Validity of Family Fixed Effects Estimates and the Long-Term Impact of Head Start." *Accessed July 31, 2017.*

Gruber, Jonathan. 1994. "The incidence of mandated maternity benefits." *The American Economic Review*, 622–641.

Han, Wen-Jui, and Jane Waldfogel. 2003. "Parental leave: The impact of recent legislation on parents' leave taking." *Demography*, 40(1): 191–200.

Han, Wen-Jui, Christopher Ruhm, and Jane Waldfogel. 2009. "Parental leave policies and parents' employment and leave-taking." *Journal of Policy Analysis and Management*, 28(1): 29–54.

Hechinger, Fred M., ed. 1966. *Pre-School Education Today: New Approaches to Teaching Three-, Four-, and Five-Year-Olds.* New York:Doubleday.

Heckman, James J., and Stefano Mosso. 2014. "The economics of human development and social mobility." *Annu. Rev. Econ.*, 6(1): 689–733.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The rate of return to the HighScope Perry Preschool Program." *Journal of Public Economics*, 94(1): 114–128.

Hoem, Jan M. 1993. "Public Policy as the Fuel of Fertility: Effects of a Policy Reform on the Pace of Childbearing in Sweden in the 1980s." *Acta Sociologica*, 36(1): 19–31.

Hofferth, Sandra L., and Sally C. Curtin. 2006. "Parental leave statutes and maternal return to work after childbirth in the United States." *Work and occupations*, 33(1): 73–105.

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics*, 6(2): 65–70.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. "Long-run impacts of childhood access to the safety net." *American Economic Review*, 106(4): 903–934.

Hoynes, Hilary, Marianne Page, and Ann Huff Stevens. 2011. "Can targeted transfers improve birth outcomes?: Evidence from the introduction of the WIC program." *Journal of Public Economics*, 95(7): 813–827.

Hudomiet, Peter. 2015. "The role of ocupation specific adaptation costs in explaining the educational gap in unemployment." Working paper.

Hunt, Joseph McVicker. 1961. *Intelligence and Experience.* New York:Ronald Press.

Johnson, Rucker C, and C. Kirabo Jackson. 2017. "Reducing Inequality Through Dynamic Complementarity: Evidence from Head Start and Public School Spending." National Bureau of Economic Research Working Paper 23489.

Kamerman, Sheila B., Alfred J. Kahn, and Paul Kingston. 1983. *Maternity policies and working women.* New York:Columbia University Press.

Kezdi, Gabor. 2004. "Robust standard error estimation in fixed-effects panel models." *Hungarian Statistical Review*, Special English Volume 9: 95–116.

Klerman, Jacob Alex, and Arleen Leibowitz. 1997. "Labor supply effects of state maternity leave legislation." *Gender and Family Issues in the Workplace. New York: Russell Sage*, 65–85.

Klerman, Jacob Alex, Arleen Leibowitz, F. Blau, and R. Ehrenberg. 1997. "Gender and Family Issues in the Workplace."

Klerman, Jacob Alex, Kelly Daley, and Alyssa Pozniak. 2012. "Family and medical leave in 2012: Technical report." Abt Associates, Cambridge, MA.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2018. "Children and gender inequality: Evidence from Denmark." National Bureau of Economic Research.

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimuller. 2019. "Child penalties across countries: Evidence and explanations." NBER Working Paper 25524.

Kline, Patrick, and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *The Quarterly Journal of Economics*, 131(4): 1795–1848.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica*, 75(1): 83–119.

Koontz, Elizabeth Duncan. 1971. "Childbirth and Child Rearing Leave: Job-Related Benefits." *New York Law Forum*, 17: 480–502.

Lalive, Rafael, Analía Schlosser, Andreas Steinhauer, and Josef Zweimüller. 2014. "Parental Leave and Mothers' Careers: The Relative Importance of Job Protection and Cash Benefits." *The Review of Economic Studies*, 81(1): 219–265.

Lalive, Rafael, and Josef Zweimüller. 2009. "How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments." *The Quarterly Journal of Economics*, 124(3): 1363–1402.

Lee, Mike, and Joni Ernst. 2019. "Sens. Ernst, Lee Put Forward Paid Parental Leave Plan That is Budget Neutral and Flexible for Parents."

Lemieux, Thomas. 2006. "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, 96(3): 461–498.

Levine, Robert A. 1970. *The Poor Ye Need Not Have With You: Lessons from the War on Poverty.* Cambridge:MIT Press.

Levitan, Sar A. 1969. *The Great Society's Poor Law: A New Approach to Poverty.* Baltimore:Johns Hopkins Press.

Levy, Helen. 2004. "Employer-sponsored disability insurance: where are the gaps in coverage?" NBER Working Paper 10382.

Loeb, Susanna. 2016. "Missing the target: We need to focus on informal care rather than preschool."

Ludwig, Jens, and Douglas L. Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *The Quarterly Journal of Economics*, 122(1): 159–208.

Mathur, Aparna, Isabel V. Sawhill, Heather Boushey, Ben Gitis, Ron Haskins, Doug Holtz-Eakin, Harry J. Holzer, Elisabeth Jacobs, Abby M. McCloskey, Angela Rachidi, Richard V. Reeves, Christopher J. Ruhm, Betsey Stevenson, and Jane Waldfogel. 2017. "Paid Family and Medical Leave."

Meyer, Bruce D., Wallace K.C. Mok, and James X. Sullivan. 2015. "Household surveys in crisis." *Journal of Economic Perspectives*, 29(4): 199–226.

Miller, Sarah, and Laura R. Wherry. 2017. "The long-term effects of early life Medicaid coverage." *Journal of Human Resources*. Forthcoming.

Mincer, Jacob A. 1974. "Schooling, Experience, and Earnings."

Mincer, Jacob, and Solomon Polachek. 1974. "Family Investments in Human Capital: Earnings of Women." *Journal of Political Economy*, 82(2, Part 2): S76–S108.

Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman. 2012. "Science faculty's subtle gender biases favor male students." *Proceedings of the National Academy of Sciences of the United States of America*, 109(41): 16474–16479.

Moulton, Brent R. 1990. "An illustration of a pitfall in estimating the effects of aggregate variables on micro units." *The Review of Economics and Statistics*, 334–338.

Mukhopadhyay, Sankar. 2012. "The Effects of the 1978 Pregnancy Discrimination Act on Female Labor Supply." *International Economic Review*, 53(4): 1133–1153.

Mulligan, Casey B., and Yona Rubinstein. 2008. "Selection, Investment, and Women's Relative Wages Over Time." *The Quarterly Journal of Economics*, 123(3): 1061–1110.

National Center for Health Statistics. 2015. "Natality Detail File, 1970-1984: [United States]." U.S. Department of Health and Human Services [producer]. Inter-university Consortium for Political and Social Research [distributor].

Neal, Derek A., and William R. Johnson. 1996. "The role of premarket factors in black-white wage differences." *Journal of Political Economy*, 104(5): 869–895.

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

Niederle, Muriel, and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, 122(3): 1067–1101.

Niederle, Muriel, and Lise Vesterlund. 2008. "Gender differences in competition." *Negotiation Journal*, 24(4): 447–463.

Niederle, Muriel, and Lise Vesterlund. 2011. "Gender and competition." *Annu. Rev. Econ.*, 3(1): 601–630.

North, A. Fredrick. 1979. "Health Services in Head Start." In *Project Head Start: A Legacy of the War on Poverty.* , ed. Edward Zigler and Jeanette Valentine. New York:Free Press.

OECD. 2018. "OECD Family Database."

OEO. 1967. "3rd Annual Report: The Tide of Progress."

OEO. 1970. "Annual Report Fiscal Years 1969-1970."

Office of Child Development (DHEW). 1970. *Project Head Start 1968: A Descriptive Report of Programs and Participants.*

Olivetti, Claudia, and Barbara Petrongolo. 2017. "The economic consequences of family policies: Lessons from a century of legislation in high-income countries." *Journal of Economic Perspectives*, 31(1): 205–230.

Pan, Jessica, Jenny Shen, Ebonya Washington, and Ilyana Kuziemko. 2018. "The 'Mommy Effect': Do women anticipate the employment effects of motherhood?"

Pei, Zhuan, Jorn-Steffen Pischke, and Hannes Schwandt. 2018. "Poorly measured confounders are more useful on the left than on the right." *Journal of Business and Economic Statistics.*

Phillips, Deborah A., Mark W. Lipsey, Kenneth A. Dodge, Ron Haskins, Daphna Bassok, Margaret R. Burchinal, Greg J. Duncan, Mark Dynarski, Katherine A. Magnuson, and Christina Weiland. 2017. "The Current State of Scientific Knowledge on Pre-Kindergarten Effects." *Accessed April 18, 2017.*

Price, Daniel N. 1986. "Cash benefits for short-term sickness: Thirty-five years of data, 1948-83." *Social Security Bulletin*, 49: 5.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How stereotypes impair women's careers in science." *Proceedings of the National Academy of Sciences*, 111(12): 4403–4408.

Rossin, Maya. 2011. "The effects of maternity leave on children's birth and infant health outcomes in the United States." *Journal of Health Economics*, 30(2): 221–239.

Rossin-Slater, Maya. 2017. "Maternity and family leave policy." National Bureau of Economic Research.

Rossin-Slater, Maya. 2018. "Maternity and family leave policy." In *The Oxford Handbook of Women and the Economy.* Oxford University Press.

Rossin-Slater, Maya, Christopher J. Ruhm, and Jane Waldfogel. 2013. "The Effects of California's Paid Family Leave Program on Mothers' Leave-Taking and Subsequent Labor Market Outcomes." *Journal of Policy Analysis and Management*, 32(2): 224–245.

Rousmaniere, Jr., James. 1977. "Chamber to press for veto of pregnancy benefit bill." *The Baltimore Sun*, A11.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2017. "Integrated Public Use Microdata Series: Version 7.0 [dataset]." University of Minnesota.

Ruhm, Christopher J. 1998. "The Economic Consequences of Parental Leave Mandates: Lessons from Europe." *The Quarterly Journal of Economics*, 113(1): 285–317.

Ruhm, Christopher J. 2000. "Parental leave and child health." *Journal of Health Economics*, 19(6): 931–960.

Sarin, Natasha. 2017. "The impact of paid leave programs on female employment outcomes." Working paper.

Sholar, Megan A. 2016. "Donald Trump and Hillary Clinton both support paid family leave. That's a breakthrough." *The Washington Post*.

Sigle-Rushton, Wendy, and Jane Waldfogel. 2007. "Motherhood and women's earnings in Anglo-American, Continental European, and Nordic Countries." *Feminist Economics*, 13(2): 55–91.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. "What Are We Weighting For?" *Journal of Human Resources*, 50(2): 301–316.

Stearns, Jenna. 2015. "The effects of paid maternity leave: Evidence from Temporary Disability Insurance." *Journal of Health Economics*, 43: 85–102.

Stoddard, Christiana, Wendy A. Stock, and Elise Hogenson. 2016. "The impact of maternity leave laws on Cesarean delivery." *BE Journal of Economic Analysis and Policy*, 16(1): 321–364.

Stuart, Bryan. 2018. "The long-run effects of recessions on education and income." Working paper.

Stuart, Bryan, Evan Taylor, and Martha Bailey. 2016. "Summary of procedure to match Numident place of birth county to GNIS places." U.S. Census Bureau 1284 Technical Memo 2.

Summers, Lawrence H. 1989. "Some simple economics of mandated benefits." *The American Economic Review*, 79(2): 177–183.

The White House Council of Economic Advisers. 2014. "The economics of paid and unpaid leave." Technical report.

Thomas, Mallika. 2018. "The Impact of Mandated Maternity Benefits on the Gender Differential in Promotions: Examining the Role of Adverse Selection." Working paper.

Thompson, Owen. 2017. "Head Start's Long-Run Impact: Evidence from the Program's Introduction." *Journal of Human Resources*, 0216–7735r1.

Timpe, Brenden. 2019. "The Long-Run Effects of America's First Paid Maternity Leave Policy."

United States Senate Committee on Labor and Public Welfare. 1964. "Economic Opportunity Act of 1964." U.S. Government Printing Office, Washington, DC.

U.S. Bureau of Labor Statistics. 2017. "National Compensation Survey: Employee Benefits in the United States, March 2017." U.S. Department of Labor.

U.S. Census Bureau. 1981. "Money income of households in the United States: 1979." U.S. Commerce Department Current Population Reports 126, Washington, DC.

U.S. House of Representatives. 1977. *Legislation to prohibit sex discrimination on the basis of pregnancy: hearing before the Subcommittee on Employment Opportunities of the Committee on Education and Labor, House of Representatives, Ninety-fifth Congress, first session, on H.R. 5055 and H.R. 6075 ... held in Washington, D.C., April 6-June 29, 1977.* Washington:U.S. Govt. Print. Off. : [For sale by the Supt. of Docs., U.S. G.P.O., Congressional Sales Office].

U.S. Senate. 1979. *Legislative history of the Pregnancy Discrimination Act of 1978, public law 95-555: prepared for the Committee on Labor and Human Resources, United States Senate.* Washington:U.S. Govt. Print. Off.

Vinovskis, Maris. 2005. *The Birth of Head Start.* The University of Chicago Press.

Waldfogel, Jane. 1998. "Understanding the "Family Gap" in Pay for Women with Children." *The Journal of Economic Perspectives*, 12(1): 137–156.

Waldfogel, Jane. 1999. "The impact of the Family and Medical Leave Act." *Journal of Policy Analysis and Management*, 281–302.

Washbrook, Elizabeth, Christopher Ruhm, Jane Waldfogel, and Wen-Jui Han. 2011. "Public Policies, Women's Employment after Childbearing, and Child Well-Being." *B.E. Journal of Economic Analysis and Policy*, 11(1).

Weinberger, Catherine J. 2014. "The Increasing Complementarity between Cognitive and Social Skills." *The Review of Economics and Statistics*, 96(5): 849–861.

Westinghouse Learning Corporation. 1969. *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development. (Executive Summary).* Clearinghouse for Federal Scientific & Technical Information, Springfield, Va.

Wisensale, Steven K. 2001. *Family Leave Policy: The Political Economy of Work and Family in America.* M.E. Sharpe, Inc.