From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2020

Essays in the Economics of Gender

Shen, Jenny

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01gm80hz38v

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Essays in the Economics of Gender

Jenny Shen

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Department of

Economics

Advisers: Janet Currie and Ilyana Kuziemko

September 2020

© Copyright by Jenny Shen, 2020.

All rights reserved.

Abstract

This collection of essays investigates the nature of gender in the economy. Chapters 1 and 2 explore educational assortative matching in family formation, starting with an overview of existing measurement tools. Using a simulation exercise, Chapter 1 shows that existing measurement tools fail to account for mechanical increases in assortativeness that arise from a changing gender gap in educational attainment. It proposes a new measure of educational assortative matching, the Perfect-Random normalization, which introduces the concept of a theoretical maximum level of assortativeness, given by the mismatch in men's and women's education distributions.

In Chapter 2, I measure trends in assortative matching among all new parents using an administrative dataset of births in the U.S. Existing work has primarily measured trends in assortative matching among married couples because of data constraints, but couples are increasingly eschewing marriage, posing a sample selection challenge. I find that sorting among all new parents has been higher than sorting among married parents since 1990.

Chapter 3, co-authored with Ilyana Kuziemko, Jessica Pan, and Ebonya Washington, examines the empirical puzzle of why women's labor force participation rates have stalled, despite women's increasing investment in human capital. We propose a hypothesis to reconcile these two trends: that when they are making key human capital decisions, women in modern cohorts underestimate the impact of motherhood on their future labor supply. Using an event-study framework, we show substantial and persistent employment effects of motherhood in U.K. and U.S. data. We then provide evidence that women do not anticipate these effects. Upon becoming parents, women (and especially more educated women) adopt more negative views toward female employment (e.g., they are more likely to say that women working hurts family life), suggesting that motherhood serves as an information shock to their beliefs. We then look at longer horizons—are young women's expectations about future labor

supply correct when they make their key educational decisions? In fact, female high school seniors are increasingly and substantially overestimating the likelihood they will be in the labor market in their thirties, a sharp reversal from previous cohorts who substantially underestimated their future labor supply.

Acknowledgements

It would not have been possible to write this dissertation without the help and support of the kind people around me. I am deeply indebted to my advisors, Ilyana Kuziemko and Janet Currie, for their expertise, guidance, and mentorship. I would like to especially thank Ilyana Kuziemko for all of the time and energy that she poured into helping me to develop into a good researcher and clear writer.

I would like to thank the faculty members and collaborators who helped me to write this dissertation. Thomas Fujiwara provided invaluable insight and encouragement throughout the writing process. Ebonya Washington was a wonderfully generous mentor and collaborator whose perceptive suggestions helped to shape much of this dissertation. I learned a great deal from Jessica Pan's deep knowledge of the economics of gender literature.

I am grateful to the Center for Health and Wellbeing and the Economics Department for being a welcoming research home for the past six years. My officemates in the CHW graduate office and especially Emily Cuddy, Jonathan Gao, and Molly Schnell provided great support and friendship throughout my Ph.D. Thank you to the CHW staff and especially Justine Conoline and Susan Rizzo for their help and for making CHW a welcoming community. I am grateful to Laura Hedden for being an invaluable resource and for answering all of my questions in great detail.

I couldn't have made it through the graduate program without the camaraderie and support of my friends (and especially the "pomodoro crew")—Tia Lee, Mauricio Matsumoto, Christine Ostrowski, Hannah Rubinton, Ben Shields, and Fernanda Sobrino. Forbes College was a fantastic place to call home for the last three years—thank you to all who made it such a welcoming place. To my New Haven friends, Molly Scott, Alejandro Sucre, Danny Stone, Karl Barber, Santiago Salazar, and Lindsey Kozak, thank you for always being down for laughs and games.

Above all, I would like to thank my family. To my partner, Paul Muir, whose endless support and sense of humor never failed to cheer me up. To my brother, who always had the time to lend a sympathetic ear. And finally to my parents, who worked tirelessly to provide my brother and me with every opportunity.

Contents

Abstract iii

Acknowledgements v

List of Tables x

List of Figures xii

1 Measuring Educational Assortative Matching 1

1.1 Introduction 1

1.2 Background 7

1.2.1 Measurement challenges 7

1.2.2 The random matching normalization 9

1.3 Simulation exercise 12

1.4 The Perfect-Random normalization 16

1.5 Results 18

1.5.1 Main results 18

1.5.2 Other sufficient statistics 19

1.5.3 Different ways of measuring education 20

1.5.4 Mechanics behind observed patterns 21

2 Trends in Educational Assortative Matching among New Parents 34

2.1 Vital Statistics 34

2.1.1 Main comparison: CPS vs. births 36

2.1.2 Missing dads 37
2.1.3 Trends by race 37
2.1.4 Trends by birth order 38
2.2 Conclusion 38

3 The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood? 54

3.1 Introduction 54
3.2 Data sources 61
3.2.1 British Household Panel Survey 62
3.2.2 U.S. panel data sources 63
3.3 Empirical strategy 65
3.4 How does motherhood affect employment? 66
3.4.1 Main results 67
3.4.2 Robustness and related results 70
3.4.3 Heterogeneous employment responses to motherhood 71
3.5 Do women anticipate the "mommy effect" in the short-run? 73
3.5.1 Effect of parenthood on views toward gender roles and female employment 74
3.5.2 Evidence from retrospective questions 79
3.5.3 Do women correctly predict the effect of motherhood on labor force participation in the short run? 80
3.6 Do young women accurately forecast their future labor supply in the longer run? 82
3.6.1 Data and variable definitions 82
3.6.2 Results 84
3.7 Why do women today underestimate the effects of motherhood? . . . 88

3.7.1 Modeling women's education and employment as a function of expected employment costs of motherhood 89
3.7.2 Predictions yielded by the model 92
3.7.3 Empirical evidence on the evolution of the costs of motherhood across cohorts 94
3.7.4 Is it harder to be a working mother today because workplaces have grown less family friendly? 96
3.8 Conclusion 99

A Mommy Effect Appendix **115**
A.1 Appendix Figures and Tables 115
A.2 Data Appendix 136
A.2.1 BHPS 136
A.2.2 NLSW 1968 139
A.2.3 NLSY 1979 143
A.2.4 PSID 147
A.2.5 Supplementary data sources asking expectations of future labor supply 151
A.3 Full model with proofs 155
A.3.1 Overview 155
A.3.2 Predictions yielded by the model 158

List of Tables

2.1 Summary statistics, by sample 52

2.2 Summary statistics of mothers and children by father's education . . 53

3.1 Summary statistics for main panel data sources 113

3.2 Relationship between college education and probability of reporting that being a parent is harder than expected 114

A.1 Summary statistics for main panel data sources, men 124

A.2 Effect of first birth on employment last week, robustness checks, BHPS and NLSW68 125

A.3 Effect of first birth on current employment, robustness checks, PSID and NLSY79 126

A.4 Effect of first birth on current employment, U.S. and U.K. men . . . 127

A.5 Impacts of second baby on current employment, women 128

A.6 Summary statistics, gender-role attitudes (BHPS) 129

A.7 Effect of motherhood on individual components of the gender-role index, controlling for a person-specific linear time trend (BHPS) 130

A.8 Summary statistics, gender-role attitudes (NLSY79) 131

A.9 Effect of first and second birth on gender-role attitudes, UK 132

A.10 Effect of first birth and second birth on gender-role attitudes, US (NLSY79) 133

A.11 Expected to work this year but wrong, robustness checks, U.K. men and women 134
A.12 Summary statistics for supplementary data sources asking about expectations of future labor supply 135

List of Figures

1.1 Educational attainment of men and women, 1950-current 23

1.2 Mother and father's education vs. probability that... 24

1.3 Perfect matching simulation exercise, visual representation 25

1.4 Trends in assortative mating according to the random-only normalization, observed matches vs. simulated (perfect) matches 26

1.5 Trends in educational assortative matching according to the random-only normalization, split by education category 27

1.6 Share of less than high school - less than high school couples in the observed data and as predicted in the perfect and random matching counterfactuals 28

1.7 Aggregate trends in educational assortative matching according to the Perfect-Random normalization vs. the Random-only normalization . 29

1.8 Trends in educational assortative mating according to the Perfect-Random normalization, by category 30

1.9 Perfect-Random normalization, other statistics 31

1.10 Perfect-Random normalization, various education categorizations . . . 32

1.11 Mechanics behind observed trends: Education correlation for observed matches vs. simulated matches 33

2.1 Trends in educational assortative matching, all parents vs. married parents only 42

2.2 Trends in educational assortative matching, missing dads = less than high school 43
2.3 Share of births / number of states reporting parental education over time 44
2.4 States in the max years sample 45
2.5 Trends in educational assortative matching as measured by the Perfect-Random normalization, simulation exercise vs. births samples, CPS . 46
2.6 Share of births where mother is unmarried and/or where father's education is missing 47
2.7 Trends in educational assortative matching for couples where the woman is black 48
2.8 Trends in educational assortative matching for couples where the woman is white 49
2.9 Trends in educational assortative matching by race of wife 50
2.10 Trends in educational assortative matching by whether first or later birth 51

3.1 Female labor force participation rates in select countries (ages 25-54) 101
3.2 Share of young adults with tertiary education, by gender 102
3.3 Event-study analysis of how probability of employment changes after motherhood, UK 103
3.4 Event-study analysis of how probability of employment changes after motherhood, US 104
3.5 Heterogeneous effects of first birth on current employment, women . 105
3.6 Event-study analysis of how gender-role attitudes change after motherhood (BHPS) 106
3.7 Event-study analysis of how gender-role attitudes change after motherhood (NLSY79) 107
3.8 Heterogeneous effects of first birth on women's gender-role attitudes 108

3.9 Event-study analysis of how expectations and reality of giving up paid work change upon motherhood, BHPS 109
3.10 Expectations and realizations of future home-making among U.S. high school seniors 110
3.11 Various proxies for employment costs of motherhood at different child ages 111
3.12 Event-study analysis of how overall job satisfaction changes upon motherhood (BHPS) 112

A.1 Average age at first birth (U.S. and U.K. women) 115
A.2 Employment over event-time for women with and without children (NLSY), assigning "placebo births" to women without children 116
A.3 Employment event-study analysis, splitting respondents by whether they mostly identify the pre- or post-periods 117
A.4 Gender-norms index over event-time for women with and without children (BHPS), assigning "placebo births" to women without children . 118
A.5 Event-study analysis of how gender-role attitudes change after fatherhood (men in the BHPS) 119
A.6 Event-study analysis of how gender-role attitudes change after motherhood (if needed, responses are reversed so that all are increasing in the pro-female-employment position) 120
A.7 Expectations and realizations of fertility (number of children) at age 38 to 42 among U.S. high school seniors 121
A.8 Evolution of various measures of breastfeeding prevalence 122
A.9 Event-study analysis of how various factors of job satisfaction change after motherhood, BHPS (Higher values indicate greater satisfaction) 123
A.10 Graphical derivation of the μ', μ'' cut-off values described in Lemma 1 163

Chapter 1

Measuring Educational Assortative Matching

1.1 Introduction

How couples form matters for dynamics of income inequality and intergenerational mobility. If high-skilled men and women tend to couple off with each other rather than with less-skilled individuals, income and other resources become more concentrated at the household level. As these couples have children, this higher concentration of resources entails higher inequality in the resources available for children's development, and, thus, potentially higher inequality in their later-life outcomes (Kremer, 1997; Fernndez and Rogerson, 2001; Fernndez *et al.*, 2005). Some have argued that marriages have become more assortative over the past 50 years (*e.g.*, Schwartz and Mare, 2005; Eika *et al.*, 2018) although there is some debate about these trends (Gihleb and Lang, 2016).

To make progress in understanding these dynamics, it is important that we accurately measure assortative matching. In this paper, I develop new tools to do so. Measuring educational assortative matching is challenging in part because men and

especially women have become more educated over time. As women have caught up to men in educational attainment, couples with similar levels of education are *mechanically* more likely to form because there are more individuals with similar levels of education to match with. When measuring educational assortative matching, researchers often aim to measure changes in sorting that occur *in excess of* changes that would naturally occur because of men's and women's changing education distributions. The reason for this distinction is that, from the perspective of studying how sorting impacts inequality and intergenerational mobility, we are generally interested in how the *matching mechanism* impacts these dynamics and thus want to net out effects that may arise from compositional shifts in education.

A common method researchers use to control for men's and women's changing education distributions is to compare observed matches to a counterfactual world where men and women match randomly (*e.g.*, Schwartz and Mare, 2005; Eika *et al.*, 2018). The logic behind this comparison is that, if we see more similarity in couples' educational attainment than we would see by random chance, we can conclude that some mechanism (preferences, search frictions) must be drawing similarly-educated individuals to each other. Moreover, the *more* similarity we see above what we would see by random chance, the *more* assortative matches are. Using this measurement paradigm, existing work has found that educational assortative matching increased between 1960 and present day (Schwartz and Mare, 2005; Mare, 1991; Blossfeld and Timm, 2003; Mare, 2016; Kalmijn, 1991; Qian and Preston, 1993).

In this paper, I show that the measured increase in educational assortative matching over the past 50 years (as measured by these relative-to-random-benchmark measures) is in fact mechanically driven by the closing of the gender gap in education. The key observation behind this result is that men and women need not have the exact same education distributions (Figure 1.1), and thus, not everyone *can* find a spouse of the same education level, even under perfectly assortative matching. As men's

and women's education distributions grow more similar, more people are able to find a spouse of the same education level, and thus, the *theoretical maximum* degree of sorting—that is, the degree of sorting that occurs if men and women match *perfectly* according to education—increases. I show that the increase in assortative matching documented in existing studies and interpreted as a rise in sorting technologies or preferences for homogamy is in fact an artifact of men's and women's educational attainment growing more similar.

Given this finding, I propose a new measure of assortative matching, which I call the Perfect-Random normalization. This measure extends and generalizes work by Liu and Lu (2006) who consider only two categories of education. Like existing studies, the Perfect-Random normalization continues to treat the random matching counterfactual as the lower bound on assortativeness but adds the perfect matching counterfactual as an upper bound.[1] The perfect matching counterfactual assumes that men and women match *perfectly* according to education—for example, college graduate men and women always match with each other to the fullest degree possible. As the gender gap in education closes, more people can match with someone of the same education level, and the assortativeness of matches in the perfect matching counterfactual—which is the *theoretical maximum* assortativeness that a society can attain—increases as well. I normalize the random matching counterfactual to equal 0 and the perfect matching counterfactual to equal 1. The interpretation of the Perfect-Random normalization is then how assortative matches are as a share of the theoretical maximum assortativeness. According to the Perfect-Random normalization, for example, college graduates have consistently matched at about 50 to 60 percent of the theoretical maximum since 1960.

[1] Because observed matches have been more assortative than the random matching counterfactual, I focus on measuring positive assortative matching. An analogous measure can be defined for negative assortative matching, where I would normalize the random matching counterfactual to be 0 and the perfect *negative* matching counterfactual to be -1 and ask where in between these two bounds the observed matches lie.

Using the Perfect-Random normalization and standard survey data sources (the Current Population Survey), I find that educational assortative matching among married couples has actually *decreased* over time, until around 2000, at which point it began to sharply increase. The year 2000 is pivotal because it is the year that men's and women's education distributions were the most equal among married individuals in the U.S. Thus, the decline in assortative matching among married couples through 2000 is due in part to the gender gap in education closing: While the *potential* for more assortative matches to occur increased, the assortativeness of observed matches did not increase to the same degree. After 2000, as women began to surpass men in educational attainment, the potential to form more assortative matches decreased, but the assortativeness of observed matches did not decline to the same degree.

Finally, I use the Perfect-Random normalization to measure educational assortative matching among all new parents using information on parental education from the Vital Statistics birth certificates dataset, an administrative dataset of all births in the U.S. Births to unmarried parents comprise over 40 percent of all births in the U.S., but existing work on assortative matching has focused on studying married or cohabiting couples. This focus is in part due to data constraints: Commonly used nationally representative datasets such as the Current Population Survey or the Decennial Census are household-level datasets, and, as such, can only identify couples who live together. Births to unmarried parents have been concentrated among couples with lower levels of education (Figure 1.2a), so we effectively exclude the lower part of the education distribution by focusing on married couples. For example, over 90 percent of advanced degree-advanced degree couples are married, but fewer than 50 percent of high school dropout-high school dropout couples are married.

In Figure 1.2b, I provide suggestive evidence that assortative matching is particularly bad for intergenerational mobility at the low end of the education distribution, a segment of the population that is particularly likely to be excluded from a sample

of married couples. In particular, a child is substantially more likely to not finish high school herself if *both* of her parents have not finished high school—that is, above and beyond the effect of *each* parent not finishing high school. Very few nationally representative datasets have information about non-cohabiting parents, and among these datasets, the Vital Statistics birth certificates data has the most historical coverage. I find that trends among *all* new parents are similar to trends among married parents only through 1990, at which point sorting among all new parents started to diverge to a higher level of sorting than among married parents.

This paper contributes to several different strands of literature. First, I contribute to the educational assortative matching literature. Existing work has used various measurement techniques to find that educational assortative matching has increased over time (*e.g.*, Schwartz and Mare, 2005; Gonalons-Pons and Schwartz, 2017; Kalmijn, 1991). Some have built upon this finding and asked whether and to what extent this measured increase in educational assortative matching has contributed to the rise in household income inequality (Greenwood *et al.*, 2014; Eika *et al.*, 2018; Schwartz, 2010; Breen and Salazar, 2011; Cancian and Reed, 1998, 1999; Karoly and Burtless, 1995). These studies differ in their conclusions about whether assortative matching has contributed meaningfully to the rise in household income inequality but mostly (all except the last 3 papers) rely on comparing observed matches to a random matching counterfactual to come to their conclusions.

Many have built upon the finding that educational assortative matching has increased over the past 50 years by exploring factors behind this rise. Some have interpreted the measured increase in assortative matching as being driven by increasing insularity of college graduates and/or the increasing isolation of high school dropouts (Schwartz and Mare, 2005; Arum *et al.*, 2008; Hou and Myles, 2008; Eika *et al.*, 2018). Others have interpreted the rise as the result of schools (rather than churches or families) playing a greater role in dating and marriage markets today

than yesterday (Kalmijn, 1991; Blossfeld and Timm, 2003), and, in particular, have hypothesized that rising homogamy is a function of the difference between age at leaving school and age at marriage (Mare, 1991; Shafer and Qian, 2010). Still others have interpreted the rise as due to the growing symmetry of men's and women's roles in marriage (Gonalons-Pons and Schwartz, 2017). See Schwartz (2013) for a review.

The Perfect-Random normalization that I propose is similar in spirit to a Gini coefficient and to methods used to measure the geographic segregation of income or race in that these other methods also compare observed phenomena to both a lower and an upper bound counterfactual distribution (*e.g.*, Logan and Parman, 2017; Reardon and Bischoff, 2011). For example, Logan and Parman (2017) bound the observed geographic racial segregation from below by assuming that all individuals randomly choose where to live (so all, e.g., MSAs have the same racial composition as the nation as a whole), and from above by assuming perfect segregation—that is, that all, e.g. white people live in a particular area, all black people another, etc.

The remainder of this paper is structured as follows: Section 2 provides further background on why measuring educational assortative matching is challenging and details some existing methods that researchers use to address those challenges. Section 3 discusses the simulation exercise that I conduct to show that a common measurement paradigm, the random matching normalization, fails to control for a changing gender gap in education. In Section 4, I discuss my proposed measure of assortative matching, the Perfect-Random Normalization, and in Section 5, I discuss the trends I find using the Perfect-Random normalization. Section 6 provides further detail on the Vital Statistics birth certificates data and discusses trends measured among all new (including unmarried) parents. Section 7 concludes and proposes future avenues for research.

1.2 Background

1.2.1 Measurement challenges

One reason researchers are interested in studying assortative matching is because of its relationship with household income inequality and intergenerational mobility. From the perspective of exploring these dynamics, education effectively serves as a proxy for human capital or earnings potential. Education is an imperfect proxy, though, because it is lumpy, and, moreover, as individuals have become more educated, the average human capital that each education level represents has likely changed. For example, the average high school dropout today is likely more negatively selected than the average high school dropout, say, in the 1940s, when a much smaller share of the population attended college.[2] Analogously, a 1960s couple with a college graduate husband and a high school graduate wife might have the same levels of human capital as, say, a college-college couple today, given the expansion in higher education in the intervening time period. The former is counted as not assortatively matched while the latter is, even though they are effectively the same couple from a measurement perspective. To the extent that we are interested in education as a sufficient statistic for earnings potential or human capital, we want to control for some of this changing selection.

A related challenge with measuring assortative matching is separating how much sorting would naturally occur (*e.g.*, if men and women matched randomly, many same-education couples would still form) from how much is due to the matching technology (*e.g.*, people's preferences for traits in partners, or search frictions that tend to draw

[2]This is a particularly lively debate in work estimatching trends in mortality among less-educated whites. As Americans have become more educated, high school dropouts are more negatively selected. The trend of rising mortality among less educated whites could be due to either: (1) Mortality staying constant at every education percentile, but increasing educational attainment means that high school dropouts are more negatively selected, or (2) Mortality is indeed rising at the low end of the education distribution, or (3) some combination of (1) and (2). See, *e.g.*, Novosad and Rafkin (2019).

certain types of people together). We can conceptualize assortative matching in many different ways—for example, a natural starting point for measuring educational assortative matching is to measure the correlation between a husband's and a wife's education or to calculate the share of couples with the same levels of education.[3] From the perspective of answering questions such as: "How much of the increase in household income inequality is due to changes in assortative matching?", researchers tend to implicitly refer to changes in assortative matching due to the matching technology, rather than say, the fact that women have become more educated, and, thus, same-educated are *mechanically* more likely to form. Measures such as the correlation between spouses' education or the share of couples with the same education include both changes that are due to men's and women's changing educational attainment as well as those due to changes in the matching technology.

From both of these reasons, men's and women's increase in educational attainment over the past 60 years or so presents a measurement challenge. Women in particular have made dramatic advances in educational attainment and in the 1990s have surpassed men in educational attainment (Goldin *et al.*, 2006a; Goldin, 2006a, 2014a; Van Bavel *et al.*, 2018). First, the increase in educational attainment means that a college degree in the 1960s represents a higher level of average human capital than a college degree today, so making apples-to-apples comparisons across time is challenging. Second, the increase in educational attainment means that there is a different portion of the educational similarity of couples that is due to the relative supply of men and women of different education levels rather than to the matching technology, and an ideal measure of assortative matching should be able to directly measure latter while controlling for the former. What researchers have relied upon

[3]Researchers use a range of other methodologies as well, depending on research questions. For example, some rely on simple measures such as the correlation between husbands' and wives' education levels, a regression of husband's years of education on wife's years of education (Greenwood *et al.*, 2014), or calculate the share of couples who share the same education level (Gihleb and Lang, 2016). Other papers have used structural methods to measure trends in assortative matching (Chiappori *et al.*, 2017; Chade and Eeckhout, 2017).

to control for changing educational attainment is to compare observed matches to a counterfactual world where men and women match randomly according to education. These comparisons take varying forms but all share the common feature of comparing to the same reference distribution of matches.

1.2.2 The random matching normalization

A common approach for controlling for changing education levels is to compare observed matches to a counterfactual world where individuals randomly match with each other (Schwartz and Mare, 2005; Eika *et al.*, 2018). Throughout this paper, I will refer to this class of measures as a "random matching" or "random-only" normalization. The logic behind these normalizations is that if there is a high prevalence of a certain characteristic—e.g., college education—then, the random matching counterfactual will also have a high prevalence of couples where both partners are college-educated, just by random chance. If the observed world has more matches of the, e.g., college-college type than this random matching counterfactual scenario, then we can conclude that some mechanism must be drawing college-educated men and women together. As men and women become more educated, the random matching counterfactual will change accordingly, and the random matching normalization measures changes in assortativeness net of what would have happened by random chance. Moreover, the *more* same-education couples we see above what we would see by random chance, the more assortative the matches are.

Eika *et al.* (2018)

Eika *et al.* (2018), for example, measure educational assortative matching using an odds measure. In particular, they calculate:

$$\text{Random matching normalization} = \frac{Pr(\text{Observed match})}{Pr(\text{Seeing same match under random matching})} \tag{1.1}$$

More formally, if E_m and E_w are random variables for a husband's vs. a wife's education, and e and e' are any two levels of education, their measure is:

$$\text{Random matching normalization} = \frac{Pr(E_m = e \text{ and } E_w = e')}{Pr(E_m = e) \cdot Pr(E_w = e')} \tag{1.2}$$

Where the denominator comes from assuming that E_m and E_w are statistically independent. They calculate aggregate trends in assortative matching by calculating a weighted average of the random matching normalization for all couples where $E_m = E_w$. A value greater than 1 indicates positive assortative matching and values less than 1 indicate negative assortative matching.

Schwartz and Mare (2005)

Schwartz and Mare (2005), as well as several other papers[4], measure educational assortative matching using the "log-linear model," a common method for measuring social stratification in sociology. The null hypothesis of the log-linear model is complete statistical independence (*i.e.*, random matching). In particular, let i be a husband's education level and j be a wife's education level, and let π_{ij} be the number of couples where the husband has education i and the wife has education j. Then,

[4] *E.g.*, Mare (1991); Kalmijn (1991); Schwartz and Mare (2012), and many others. See Agresti (2003).

statistical independence means that:

$$\pi_{ij} = A \cdot \alpha_i \cdot \beta_j \tag{1.3}$$

Where A, α_i, and β_j are parameters to be fit. Usually, researchers take logs of the above equation, so that:

$$\log(\pi_{ij}) = A' + \alpha'_i + \beta'_j \tag{1.4}$$

Where again, A', α'_i, and β'_j are parameters to be fit, but they can now be fit using a linear regression. An additional residual term added to Equation 1.4 captures how assortative matches are, net of what would be predicted by random matching. For example, to measure aggregate trends, Schwartz and Mare (2005) add in a dummy variable equal to 1 if a husband's and a wife's education is the same and zero otherwise (the "homogamy model," equation (2)). To isolate particular parts of the education distribution that are driving the aggregate trends, they estimate a "crossings model", where they add in several dummy variables that estimate the relative difficulty of crossing each education boundary (Schwartz and Mare (2005), equation (3)).

Greenwood *et al.* (2014)

Greenwood *et al.* (2014) compare observed matches to random matching in their analysis of whether assortative matching has contributed to the rise in household income inequality.[5] In particular, to control for changing educational attainment over time, they calculate Gini coefficients using observed matches and compare them to Gini coefficients calculated assuming that men and women match randomly.

[5]They do not use the random matching normalization when measuring trends in assortative matching, though.

1.3 Simulation exercise

Researchers claim that the random matching normalization fully controls for the mechanical changes in the educational similarity of spouses that arise from changes in men's and women's educational attainment. As a test of that claim, I run a simulation exercise where I assume perfect assortative matching in every year—for example, college graduate men and women always match with each other to the fullest degree possible, and "leftover" college graduate men match with some college women.[6] While matching by construction never changes in my simulation exercise, I allow the education distributions of men and women to change each year according to the data. Therefore, because the only changes occuring in my simulation exercise are the changing education distributions of men and women, the only changes in the educational similarity of spouses are *mechanical*, and, thus, the random matching normalization should not measure any changes in educational assortative matching over time (that is, it should remain constant). What I will show, however, is that the trends I measure in this perfect matching simulation exercise actually *replicate* the measured trends found in existing work.

The specifics of my simulation exercise are as follows (See Figure 1.3 for a visual representation): I assume that everyone is characterized by a latent and continuous (scalar) characteristic which I call human capital. In each year t, an equal number of men and women are separately ranked according to their human capital, so that there is a "top" man and a "top" woman. Every year, everyone forms a match and matching is always perfect: the "top" man and "top" woman always match, as do the second-ranked man and woman, and so on. While matching by construction never changes in my simulation exercise, how men and women of different human capital ranks sort into different education categories does change. In particular, the top block of men and women are always assigned to be college graduates, the next block are

[6] Or vice versa, when there are more college graduate women than men.

always assigned to have some college education but no degree, and the bottom-most block is always assigned to have less than a high school degree. I allow the sizes of these blocks to differ by gender and to change each year according to the observed education distributions of men and women in my data.

I examine the random matching normalization as implemented by Eika *et al.* (2018) because their implementation is the most straightforward, but similar arguments apply to other analyses that compare observed matches to the random matching counterfactual.[7] In order to make direct comparisons, I follow Eika *et al.* (2018) in sample definition and how I measure education. In particular, I examine married couples in the Current Population Survey where at least one spouse is aged 26-60. I measure education using 4 categories: 1 = Less than high school (<12 years), 2 = high school graduate (12 years), 3 = some college but no degree (13-15 years), and 4 = college graduate (16+ years).

I plot the measured assortativeness of matches in my perfect matching simulation exercise in Figure 1.4 together with the measured trends in assortative matching found in Eika *et al.* (2018). Because matching by construction never changes in my simulation exercise, an ideal statistic to measure assortative matching should also remain unchanged, even as men's and women's education distributions change. What I find, however, is that the assortativeness of the perfectly matched couples in my simulation exercise has *increased* over time according to the random-only normalization (the dashed red line). Indeed, the assortativeness of the simulated perfect matches has increased in parallel with the assortativeness of observed matches (the solid blue line).

The intuition behind this result is that as women have caught up to men in educational attainment—that is, as the education distributions of men and women have

[7]Unfortunately, because my perfect matching simulation exercise results in many cells having no matches—for example, there are never any matches between a high school dropout and a college graduate in my perfect matching simulation exercise—I cannot measure trends using the log-linear model used by Schwartz and Mare (2005). Doing so would involve taking logs of zeros.

grown more similar—men and women have become more able to find spouses of the same education level. Consider, in the 1960s, when about ten percent of men and six percent of women were college graduates. Assuming everyone matches, the maximum share of college-college couples that can form is $\min(10, 6) = 6$ percent. In the perfect matching scenario, the remaining 4 percent of college-educated men would match with women with some college education but no degree. As the education distributions of men and women grew more similar in the 1990s, all college-educated men could find a match with a college-educated woman, so no "leftover" college-educated men matched with "some college" women. Thus, the measured assortativeness of matches in the perfect matching simulation increased relative to the 1960s when education distributions were more dissimilar, even though the underlying matching mechanism remained the same.

As an additional check, I re-run my simulation exercise, but I assume no gender gap in education. That is, I maintain the assumption that everyone matches perfectly according to human capital each year (the top man and woman always match, etc.), and continue to assign men and women to education categories based on their human capital rank. Instead of allowing the education distributions to differ by gender, however, I assume in each year that men and women have the same education distribution, given by the education distribution of the pooled sample of men and women for each year. I plot the assortativeness of matches in this alternative version of the perfect matching simulation exercise also in Figure 1.4 (the dotted green line) and find that the random matching normalization is constant over time in this alternative perfect matching simulation exercise, even though men's and women's education is still increasing over time.

At the education-category level, I also find that the assortativeness of the perfectly matched couples in my simulation exercise (allowing education distributions to differ by gender) has changed over time according to the random-only normalization, even

though matching by construction does not change (Figure 1.5b). In fact, the measured trends in my simulation exercise closely resemble measured trends in the "real world" (Figure 1.5a). For example, according to the random-only normalization, "real world" sorting among college graduates has declined over time. In my simulation exercise, sorting among college graduates has declined as well. Similarly, "real world" sorting among high school dropouts has increased over time; in my simulation exercise, I find the same trend.

If the education levels of men and women had increased in lockstep—that is, if men and women always had the same education distributions, and they changed in the exact same way—then the random-only normalization would suffce for controlling for changing educational attainment. However, because the gap in educational attainment changed in addition to men and women becoming more educated, the random-only normalization fails to fully control for mechanical increases in assortativeness arising from changing educational attainment.

Existing work on the log-linear method has hinted at this issue of the random matching normalization not being able to control for a changing *gap* in the marginal distributions but has never fully addressed it (Logan, 1996; Hellevik, 2007; Xie and Killewald, 2013).[8] The odds measure as calculated by Eika *et al.* (2018) has been used extensively in intergenerational mobility research and is usually calculated using a contingency table, which is essentially a cross-tabulation of husband's education along one dimension and wife's education along the other. Viewed another way, a contingency table is a square matrix. When an entire row or an entire column in the matrix is multiplied by a constant, the odds measure for cells in that row or column do not change. This "proportional invariance" property of the odds measure has

[8]The log-linear method is closely related to the odds measure as calculated by Eika *et al.* (2018). In particular, assortativeness in the log-linear model is the log of the odds calculated by Eika *et al.* (2018).

in general been overinterpreted to mean that the odds measure is invariant to *any* change in the marginal distributions.

1.4 The Perfect-Random normalization

Given the results of my simulation exercise, I propose a new measure of assortative matching, which I call the Perfect-Random normalization. I define the Perfect-Random normalization as follows: I normalize the random matching counterfactual to equal 0 and the perfect matching counterfactual to equal 1 and ask where in between these two bounds the observed matches lie.[9] More concretely, my normalization is:

$$\text{Perfect-Random Normalization} = \frac{\text{Observed} - \text{Random}}{\text{Perfect} - \text{Random}} \tag{1.5}$$

Note that if observed matches coincide with perfect matching (*i.e.*, Observed = Perfect), this normalization equals 1, and, conversely, if observed matches coincide with random matching (Observed = Random), this normalization equals zero.

The Perfect-Random Normalization is easiest to understand visually. In Figure 1.6, I plot the share of high school dropout-high school dropout couples as predicted under the perfect and random matching counterfactuals (the red dashed line and green dotted lines, respectively) and as observed in the real world (the solid blue line). As can be seen from this figure, the three lines move in parallel, with random matching effectively serving as a lower bound and perfect matching effectively serving as an upper bound.[10] The numerator of the perfect-random normalization (Figure 1.6a) is

[9] If observed matches are *below* the random matching counterfactual, then I treat random matching as the upper bound (normalized to be equal to 0) and simulate perfectly *negative* assortative matching—that is the "top" man matches with the N-th ranked woman, 2 matches with $N-1$, etc. I then treat the perfectly negative assortative matching counterfactual is then the lower bound, normalized to be -1, and the Perfect-Random normalization (for negative assortative matching) asks where in between these two bounds the observed matches lie.

[10] During the period observed in my data (1962-present), assortative matching by education has always been positive—that is, there have always been more same-education couples than as predicted under random matching. However, one could also imagine that perfect *anti*-matching serving as

the distance between the observed and random matching lines, and the denominator of the perfect-random normalization (Figure 1.6b) is the distance between the perfect and random matching lines.

How I implement the perfect-random normalization is category-by-category.[11] More formally, let E_m and E_w be random variables for a man's and a woman's education level and let e and e' be any two levels of education. Denote $p_{sim}(e, e')$ where $sim \in \{\text{Perfect}, \text{Random}, \text{Observed}\}$ as the probability of observing a match with $E_m = e$ and $E_w = e'$ under each matching scenario. Then the Perfect-Random normalization is:

$$s(e, e') = \frac{p_{\text{observed}}(e, e') - p_{\text{random}}(e, e')}{p_{\text{perfect}}(e, e') - p_{\text{random}}(e, e')} \tag{1.6}$$

The random matching simulation is random in a statistical sense—that is, I assume that E_m and E_w are independent random variables, so $p_{\text{random}}(e, e') = \Pr(E_m = e) \cdot \Pr(E_w = e')$.[12] To calculate aggregate trends in assortative matching, I calculate a weighted average of the perfect-random normalization calculated for when a couple has the same education level and the weights are the share of observed matches in each of these cells, following the spirit of Eika et al. (2018).

a true lower bound. The way I think about the Perfect-Random Normalization is that whether the observed matches lie above or below the random matching counterfactual determines the *sign* of assortative matching—that is, whether it is positive or negative assortative matching. *How* assortative the matches are—that is, the *magnitude* of assortative matching—is then the distance between the random matching line and either the perfect matching or perfect anti-matching lines, depending on whether there is positive or negative assortative matching.

[11] Later, in Section 5.2, I also calculate the Perfect-Random normalization in aggregate—e.g., instead of using category-by-category probabilities, I calculate the correlation between spouses' education under each reference distribution and plug these values into the same perfect-random normalization.

[12] It is difficult to write a clean formula for $p_{\text{perfect}}(e, e')$ because it depends on how "leftover" men and women match.

1.5 Results

1.5.1 Main results

Figure 1.7 plots aggregate trends in assortative matching as measured by the Perfect-Random normalization. In contrast to existing work that has found that assortative matching has increased over time, I find that educational assortative matching has declined over time, until around 2000, at which point it began to increase again. Again, the Perfect-Random normalization controls for the changing gender *gap* in educational attainment. The downward trend through 2000 is driven in part by increasing potential for more assortative couples to form; observed assortativeness did not increase at the same rate. The subsequent increase in assortative matching after 2000 driven by the gender gap re-emerging as women have surpassed men in educational attainment.

Looking at component parts in Figure 1.8, I find that college graduates and high school dropouts have historically been the most assortative groups, sorting at about 50-60 percent of the theoretical maximum since the 1960s. Up until 2000, college graduates were the most assortative, but around 2000, high school dropouts surpassed college graduates as the most assortative group, which is consistent with a story that high school dropouts are increasingly isolated from those with higher levels of education (a common hypothesis in research measuring the mortality rates of non-hispanic whites with lower levels of education—e.g., Novosad and Rafkin (2019)). In contrast, sorting among high school graduates and some college has been lower, with sorting among those with some college education but no degree falling to about 20 percent of the theoretical maximum in recent years (a.k.a., it has been pretty close to random).

1.5.2 Other sufficient statistics

I implemented the Perfect-Random normalization category by category and then calculated a weighted average, but one could calculate the Perfect-Random normalization in other ways as well. In particular, we can also calculate the Perfect-Random normalization using the correlation between a husband's and a wife's education as a sufficient statistic, or, alternatively, we could use the share of couples with the same level of education as a sufficient statistic.

Let $corr_{\text{sim}}$ with $\text{sim} \in \{\text{Observed}, \text{Perfect}, \text{Random}\}$ denote the correlation between a husband's and a wife's education in either in the observed data, or alternatively in the perfect matching or random matching counterfactuals. Then the Perfect-Random normalization calculated using the correlation is:

$$\text{Perfect-Random}_{\text{corr}} = \frac{\text{corr}_{\text{observed}} - \cancelto{cov(E_m,E_w)=0}{\text{corr}_{\text{random}}}}{\text{corr}_{\text{perfect}} - \cancelto{cov(E_m,E_w)=0}{\text{corr}_{\text{random}}}} = \frac{\text{corr}_{\text{observed}}}{\text{corr}_{\text{perfect}}} \tag{1.7}$$

Note that because we assume statistical independence in the random matching counterfactual, then $cov(E_m, E_w) = 0$ (where E_m and E_w denote a man's and a woman's education).

Similarly, for the share of couples with the same level of education, we can calculate:

$$\text{Perfect-Random}_{\text{share same}} = \frac{\mathbb{E}[\mathbb{1}[\text{Same}]_{\text{observed}}] - \mathbb{E}[\mathbb{1}[\text{Same}]_{\text{random}}]}{\mathbb{E}[\mathbb{1}[\text{Same}]_{\text{perfect}}] - \mathbb{E}[\mathbb{1}[\text{Same}]_{\text{random}}]} \tag{1.8}$$

Figure 1.9 plots trends in educational assortative matching as measured by the Perfect-Random normalization using these various sufficient statistics. As can be seen from the figure, trends across these various implementations of the Perfect-Random normalization largely agree, although assortative matching as measured by correlation is higher than as measured by share same (or by using the weighted average). The correlation takes into account the *order* of the education categories—e.g.,

that "some college" is closer to "college graduate" than "high school dropout" is to "college graduate"—whereas the other two sufficient statistics do not.

1.5.3 Different ways of measuring education

One critique of existing work measuring educational assortative matching is that trends in assortative matching are sensitive to how education is measured (Gihleb and Lang, 2016). In Figure 1.10, I calculate the Perfect-Random normalization using different categorizations of education. In particular, I examine the following different education categorizations, following the categorizations defined by (Gihleb and Lang, 2016):

- 2 categories: High school grad and below vs. some college and above (a.k.a., "skilled" vs. "unskilled")
- 4 categories (same as before): Less than high school (<12 years), high school grad (12 years), some college (13-15 years), college graduate (16+ years)
- 5 categories: 4 categories, but splits out college graduate into college graduate (16 years) and advanced degree (>16 years)
- 6 categories: 5 categories, but splits out less than high school (<12 years) into middle school ($\leq$ 8 years) and high school (9-11 years)

To measure trends in assortative matching, I rely on the weighted average of the Perfect-Random normalization calculated for each combination of husband's/wife's education (equation (4)). Figure 1.10 plots the results. While the trends are not exactly the same (nor we would expect them to be, as they measure different phenomena), the trends are qualitatively similar: Assortative matching, across all methods of measuring education, still declines through 2000 and rises thereafter. What differs across the different methods of measuring education is how much of a dip

occurs. In particular, splitting out advanced degree from college graduate leads to a *bigger* drop in assortative matching through 2000 than combining those two categories (i.e., by measuring education using 5 vs. 4 categories). Assortative matching as measured using 2 categories (i.e., "skilled" vs. "unskilled") leads to a finding that assortative matching overall is much higher than as measured using more granular categorizations.

1.5.4 Mechanics behind observed patterns

A common pattern across the various education categorizations is declining sorting through 2000 followed by increasing sorting thereafter. We can see the drivers behind these trends by separately plotting the correlations between spousal education under perfect matching, observed matching, and random matching. By construction, the correlation between spousal education under random matching will be essentially zero[13]. Thus, we can focus our attention on the trends in the correlation of spousal education under the observed and perfect matching scenarios, plotted in Figure 1.11a. Recall that the perfect + random normalization is:

$$\frac{\text{Observed} - \text{Random}}{\text{Perfect} - \text{Random}}$$

This normalization, when using the correlation as the sufficient statistic, essentially simplifies to:

$$\frac{\text{Observed}}{\text{Perfect}}$$

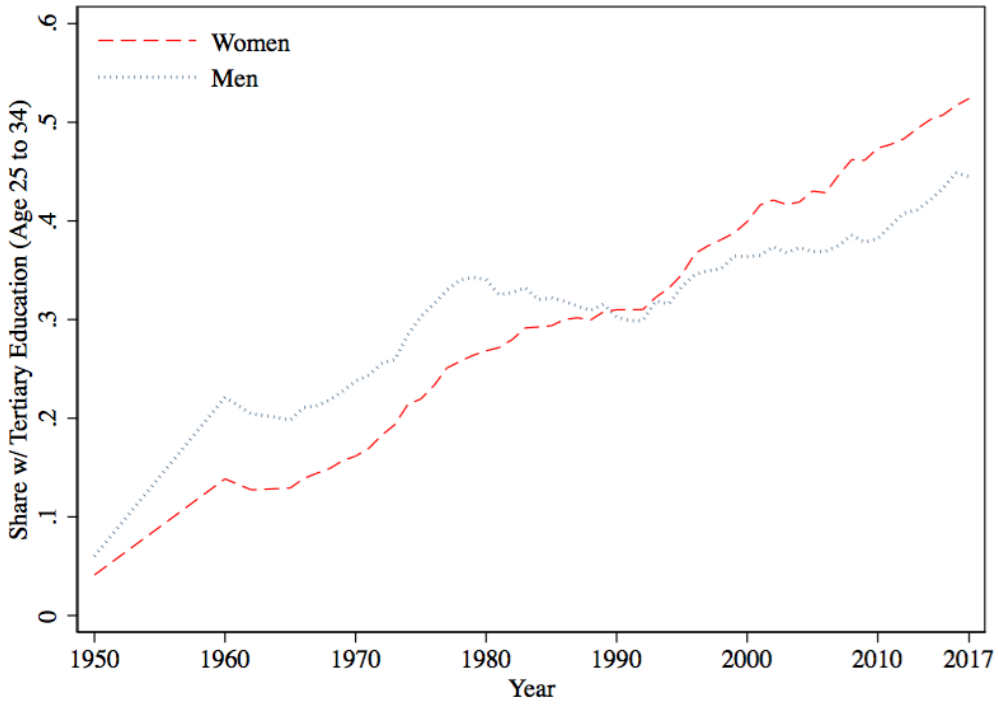
Recall that the perfect matching scenario is the maximum value that the correlation in spousal education can reach, given that the most educated men and women

[13]It will not be exactly zero because men and women have different education distributions, so not all college-educated women can be matched with college-educated men, since there are more college-educated women than men. This type of "structural" mismatch will cause the correlation of spousal education under random matching to deviate from zero.

match with each other. Figure 1.11b shows why the year 2000 is such a significant year: It is the year in my sample when women surpassed men in attending college, aka the year when men and women had the most equal education distributions. Because the education distributions were so equal in that year, the correlation under perfect matching is as close as possible to 1. In the years after 2000, the correlation of spousal education under perfect matching declines as women become increasingly more educated than men, and, therefore, more college-educated women must "marry down." In the years prior to 2000, the reverse is occurring: the maximum possible correlation of spousal education increases as women catch up to men in educational attainment.

Figures

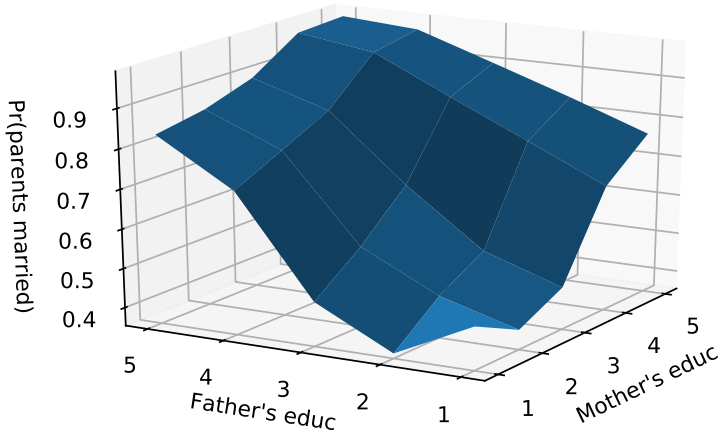
Figure 1.1: Educational attainment of men and women, 1950-current



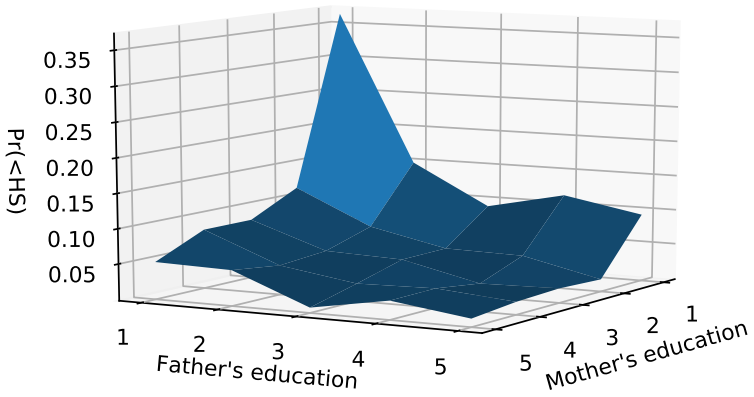
Source: Decennial Census (1950, 1960), Current Population Survey (1962-2017)
Notes: Sample is all men and women aged 25 to 34, and tertiary education is defined as an Associate's degree or above.

Figure 1.2: Mother and father's education vs. probability that...

(a) Parents are married



(b) Child does not finish high school



Source: Vital Statistics Natality, 2016 (left); General Social Survey (right)
Notes: Education categories: 1 = <12 years; 2 = 12 years; 3 = 13-15 years; 4 = 16 years; 5 = >16 years

Figure 1.3: Perfect matching simulation exercise, visual representation

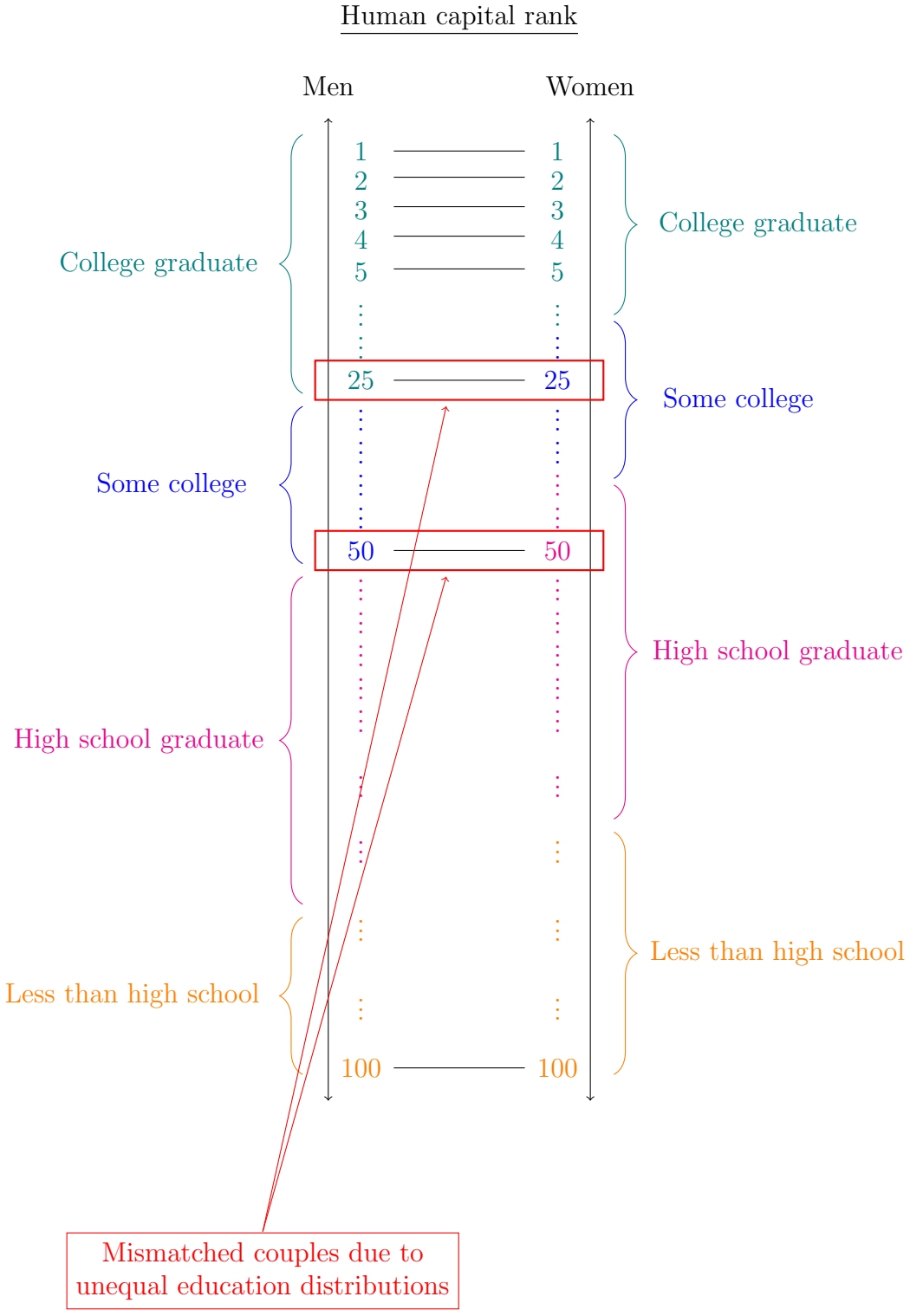
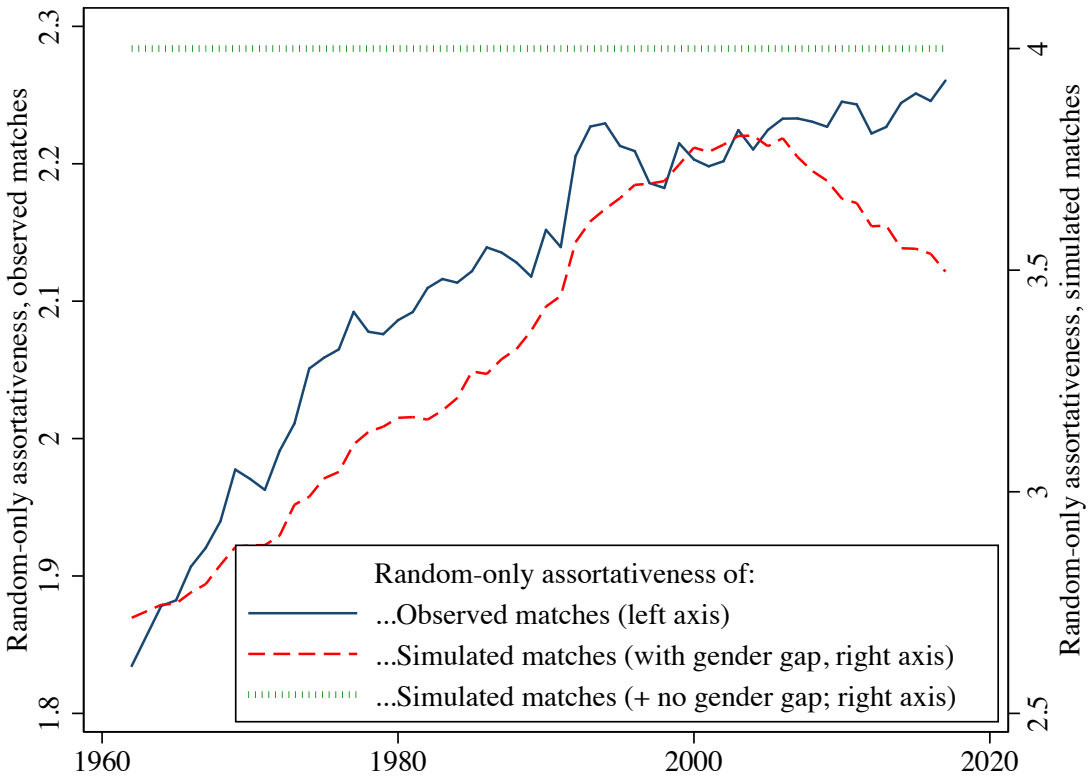
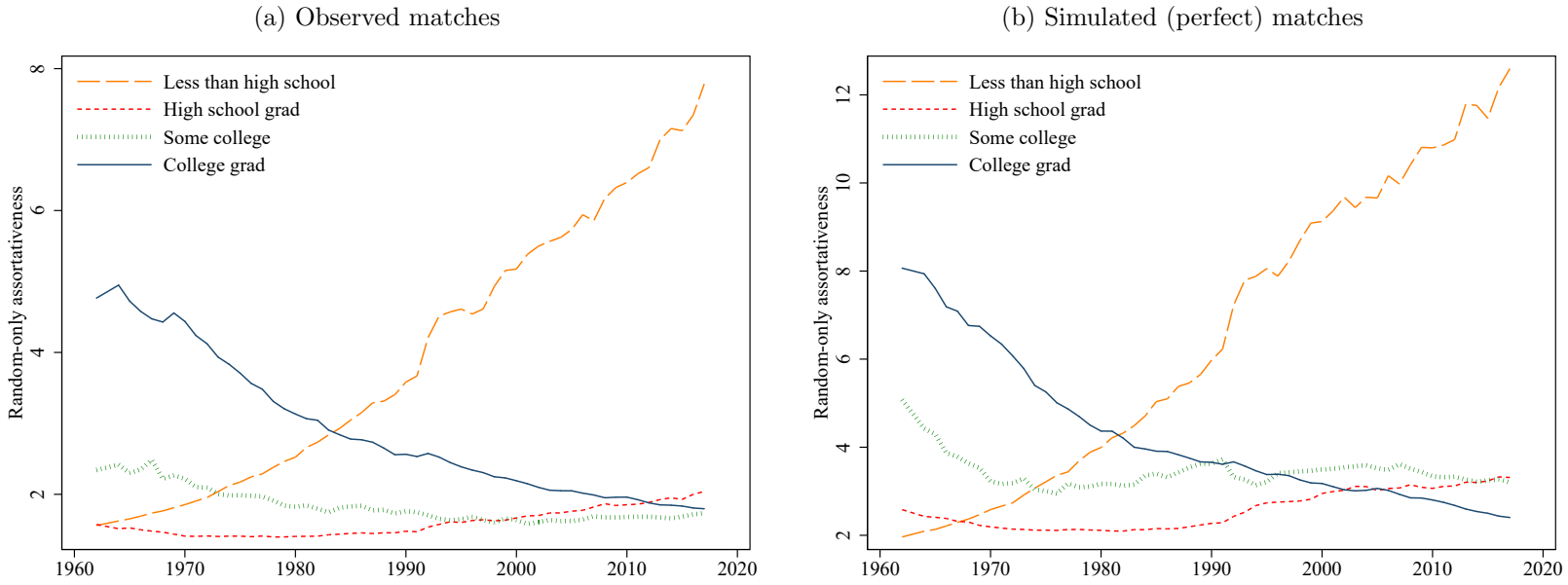


Figure 1.4: Trends in assortative mating according to the random-only normalization, observed matches vs. simulated (perfect) matches



Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling). I calculate the random-only normalization according to the normalization proposed by Eika et al. (2018). That is, for each education category, I calculate the odds (relative to random matching) of seeing a couple where both partners have that education level and then calculate a weighted average of these odds, where the weights are the share of same-education matches that fall in that cell.

Figure 1.5: Trends in educational assortative matching according to the random-only normalization, split by education category

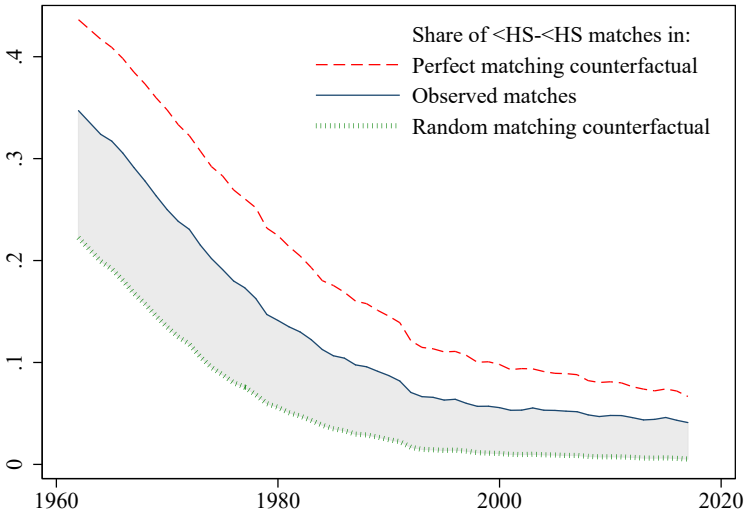


Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling). I calculate the random-only normalization according to the normalization proposed by Eika et al. (2018). That is, for each education category, I calculate the odds (relative to random matching) of seeing a couple where both partners have that education level.

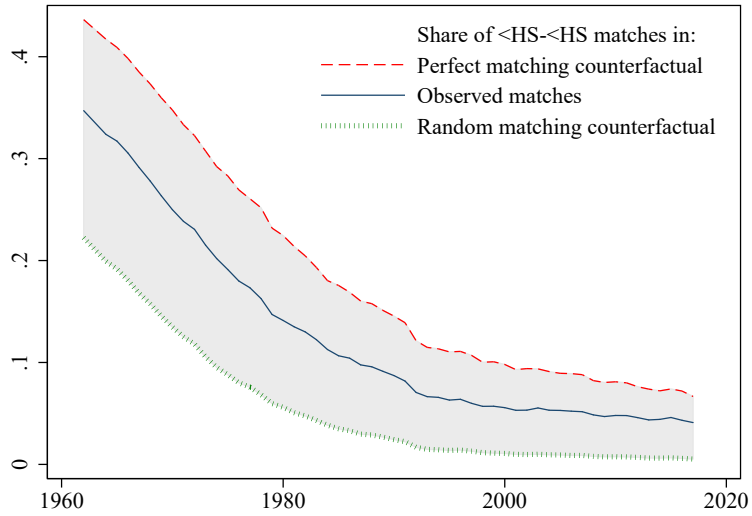
Figure 1.6: Share of less than high school - less than high school couples in the observed data and as predicted in the perfect and random matching counterfactuals

$$\text{Perfect-Random Normalization} = \frac{\text{Observed} - \text{Random}}{\text{Perfect} - \text{Random}}$$

(a) Numerator = Observed − Random



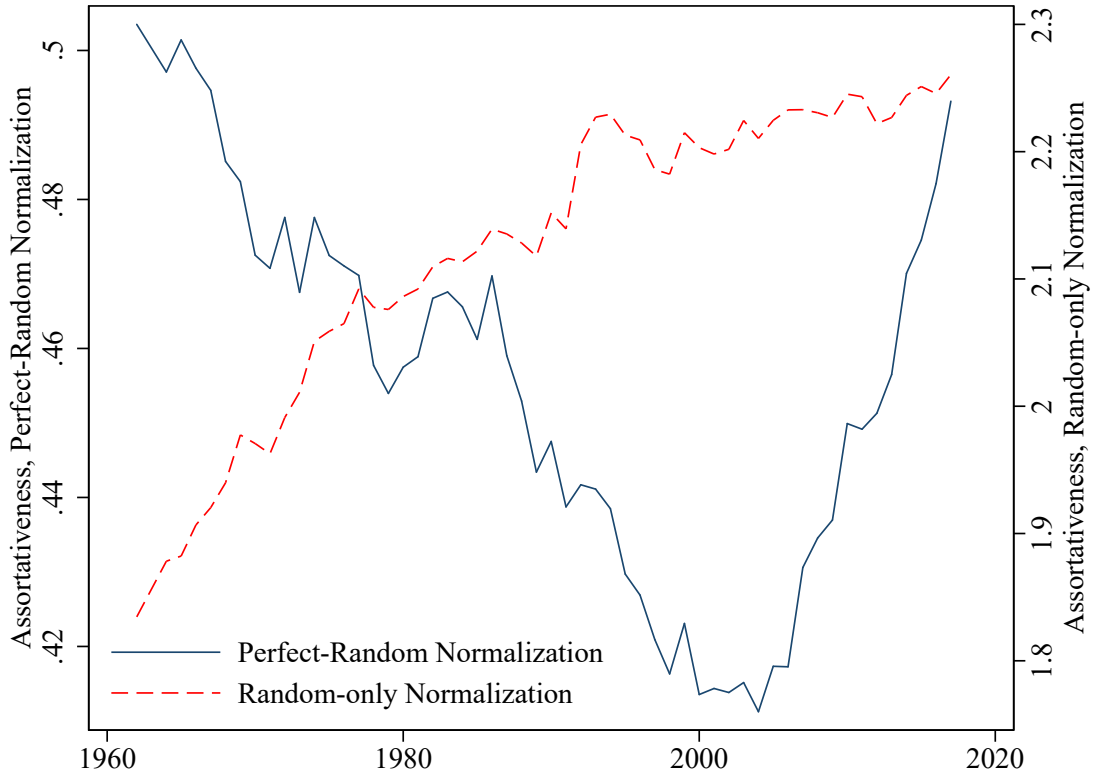
(b) Denominator = Perfect − Random



Source: Current Population Survey

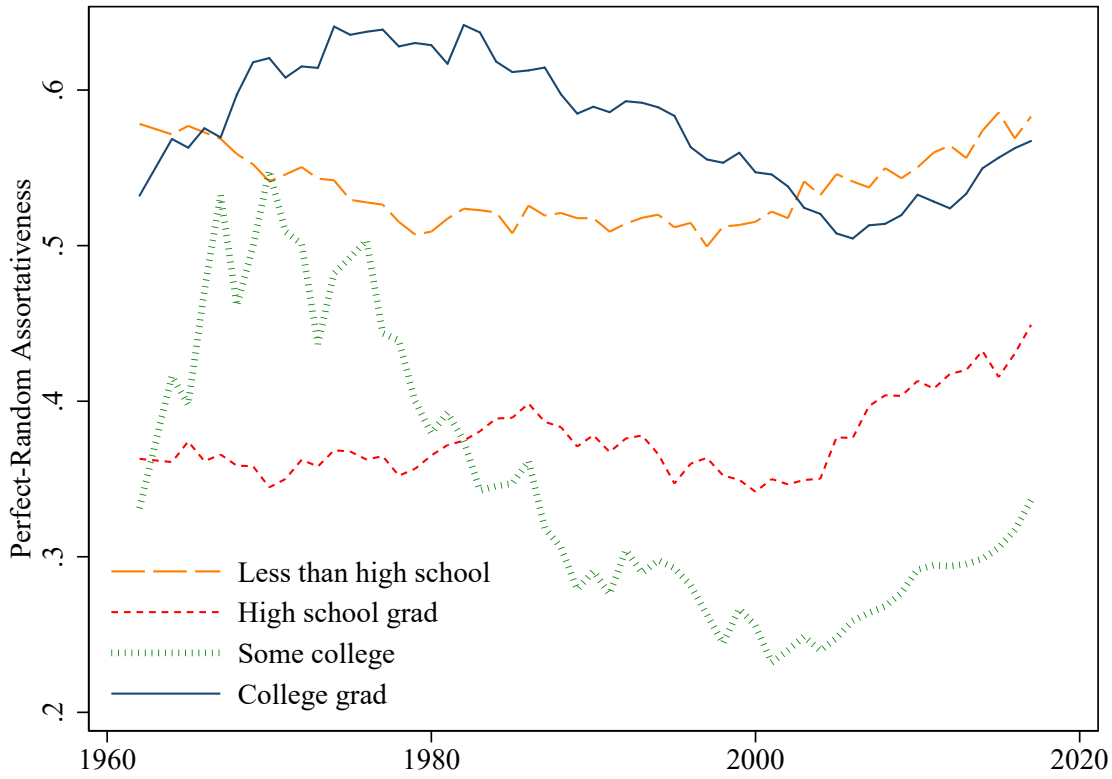
Notes: Sample is all married couples where at least one spouse is between the ages of 26-60, following Eika et al. (2018). Each line plots the share of couples where both partners have 12 years of education / is a high school graduate. The random matching counterfactual assumes that men and women match randomly, so that the probability of seeing a high school-high school is equal to Pr(a woman is a high school grad) · Pr(a man is a high school grad). The perfect matching counterfactual assumes that men and women match perfectly according to education to the maximum degree possible. So, for example, if there are more college graduate men than women, all of the college graduate women match with college graduate men, and the "leftover" college graduate men match with some college women.

Figure 1.7: Aggregate trends in educational assortative matching according to the Perfect-Random normalization vs. the Random-only normalization



Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling).

Figure 1.8: Trends in educational assortative mating according to the Perfect-Random normalization, by category



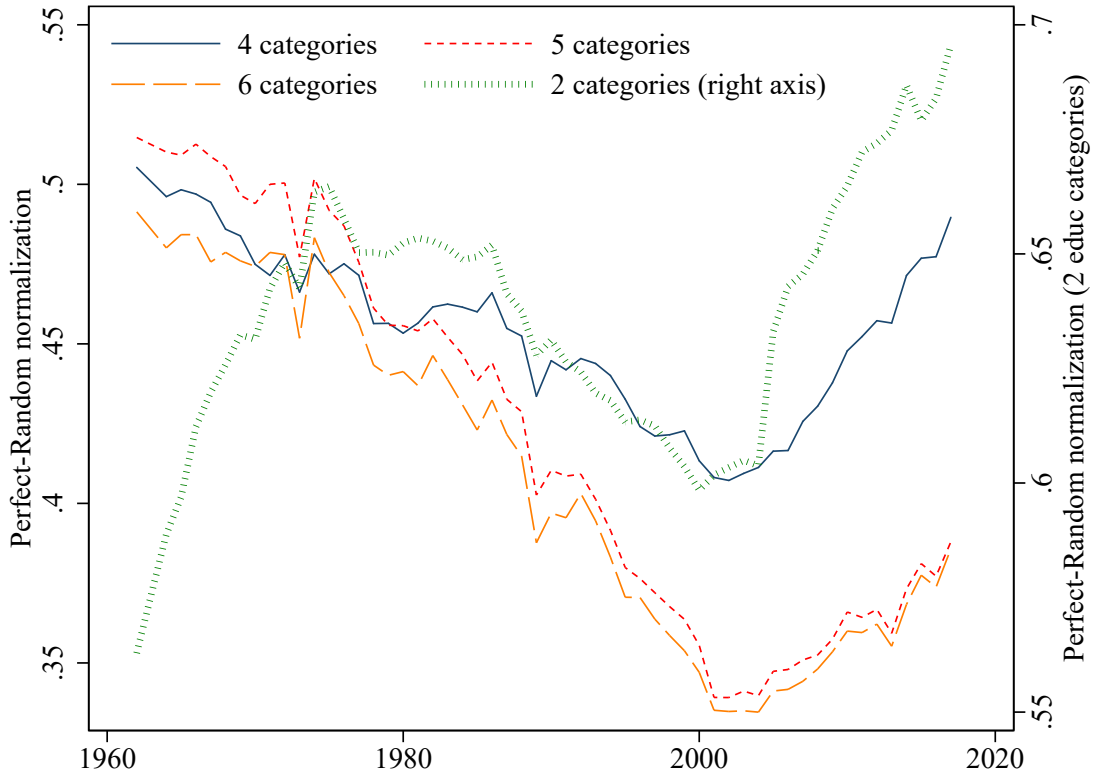
Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling).

Figure 1.9: Perfect-Random normalization, other statistics



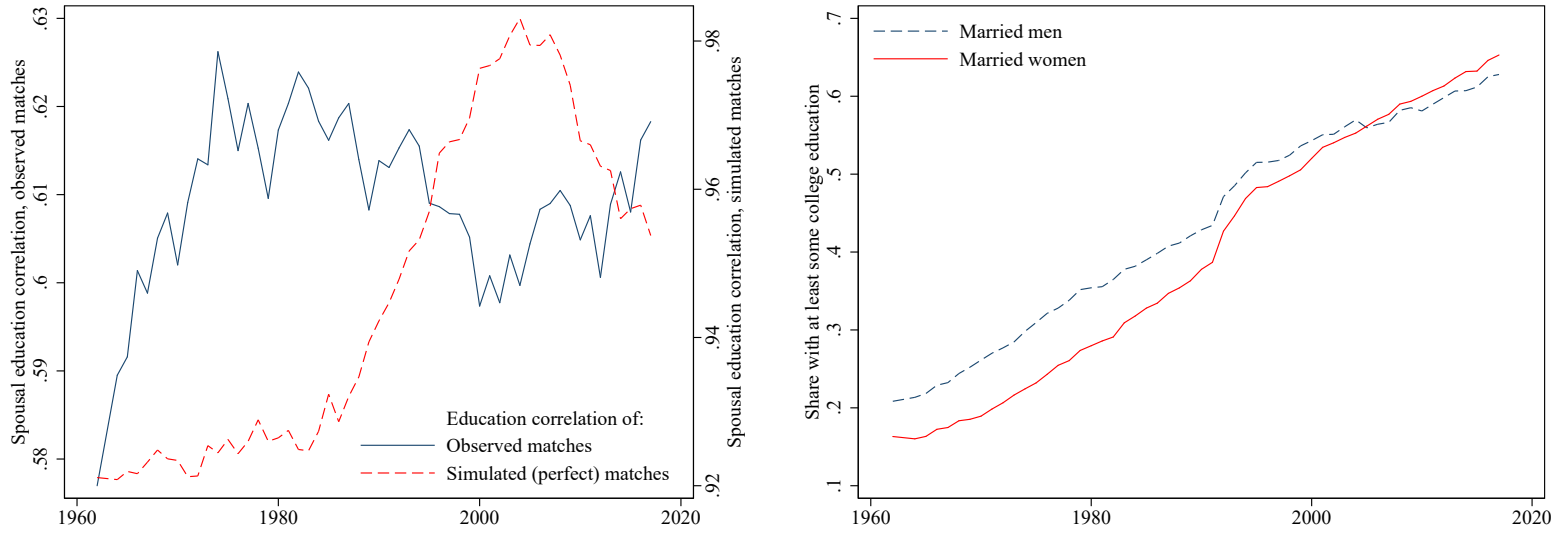
Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling).

Figure 1.10: Perfect-Random normalization, various education categorizations



Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. [talk about different education categorizations] Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling).

Figure 1.11: Mechanics behind observed trends: Education correlation for observed matches vs. simulated matches



Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling).

Chapter 2

Trends in Educational Assortative Matching among New Parents

2.1 Vital Statistics

So far, we have focused on studying married or cohabiting couples because of data constraints: Nationally representative datasets such as the Current Population Survey and the Decennial Census are household-level datasets and, as such, can only identify couples who live together. Studying only married couples excludes a large relevant swath of the population when studying questions relating to intergenerational mobility. In particular, births to unmarried parents rose to 40 percent in recent years, and, moreover, have been concentrated among couples with lower (less than college) levels of education. Therefore, studying only married parents in the intergenerational mobility context effectively excludes couples at the bottom of the education distribution. Very few nationally representative datasets have information about non-cohabiting parents, and among these datasets, the vital statistics birth certificates data has the most historical coverage.

In my analysis, I use information on parental education recorded on administrative birth certificates records, which allows for studying trends in assortative matching among unmarried, non-cohabiting parents in addition to the standard married couples sample. Beginning with the 1968 Standard Birth Certificate form, states in the U.S. began to collect information on parents' education. In addition to parental education, the birth certificates data also include other demographic information on parents such as mother's and father's race, marital status, mother's and father's age, and mother's county of residence, as well as information on the child such as birthweight and gestational age.

Unfortunately, because state-level vital statistics offices are responsible for collecting data, states differ in when they began reporting parental education. In 1969, for example, only 36 states reported parents' education, but, as more states adopted the 1968 Standard Birth Certificate Form, that number gradually increased. Another issue with the birth certificates data is that the Centers for Disease Control, the federal agency responsible for collecting data from individual states, did not collect information on father's education between 1995 and 2008, so I am unable to speak to trends in assortative matching during that time period. Finally, individual birth certificates may be missing father's education because a mother may opt to not include it.

To address selection of states into and out of education reporting, I examine two primary samples, which I call a "max years" sample and a "max states" sample. The max years sample consists of a panel of states that consistently report education over time.[1] For ease of computation, I examine a 1% random sample from each year of the max years sample (1969-2016). Conversely, the max states sample maximizes geographic coverage at the expense of some of the earlier years when fewer states

[1]The specific states that comprise the max years sample are those states with data available for all years. 19 states meet this requirement, and they are: Colorado, Iowa, Indiana, Kansas, Kentucky, Michigan, Montana, North Dakota, Nebraska, New Hampshire, Nevada, Oklahoma, South Carolina, South Dakota, Tennessee, Utah, Vermont, and Wyoming. See Figure 2.4 for data availability and reporting of parental education.

reported parental education.[2] I include all births in the max states sample since the max states sample consists of a relatively small number of years.

For comparison, I also examine trends in the Current Population Survey. To create a comparable sample to the births dataset, I examine married couples in the CPS where at least one spouse is between the ages of 25-34 and the household contains at least one own child under the age of 5. Table 2.1 reports summary statistics for the CPS sample as well as three additional samples: (1) A 1% random sample of all births data available, (2) the max years sample (a subset 1% random sample), and (3) the max states sample (but with *all* births in the years chosen. The main contrast between these samples is the rates of marriage: Whereas all women are married in the CPS sample, roughly 70 percent are married in the various births samples. As such, these samples are less white (particularly the fathers), and less educated (for example, there are more mothers and fathers who haven't finished high school). While the CPS and max years 1% sample are roughly the same size, the max states sample contains many more couples. Reassuringly, the samples look pretty similar between the max years and max states samples, although the max years sample is slightly whiter.

2.1.1 Main comparison: CPS vs. births

Figure 2.1 plots trends in educational assortative matching for parents in the births samples as well as the parents in the CPS. The trends across these samples are remarkably similar, although sorting in the max states sample is slightly higher starting around 1990. Restricting the CPS sample to couples where at least one spouse is aged 25-34 (versus 26-60 for earlier results) does not change trends much through 2000 but

[2]The years in the "max states" sample are: 1978, 1989, 1994, and 2016. In all years, only a handful of states did not report parents' education. In 1978, these were Washington, Texas, and New Mexico. In 1989, these were Washington and New York (excluding New York City, which did report parental education) In 1994, all states except Washington, D.C. reported. In 2016, all states reported parental education.

does result in a steeper rise in assortative matching since 2000 (Figure 2.5). Now, the rise in assortative matching after 2000 is steeper than the decline through 2000.

2.1.2 Missing dads

Roughly 15 to 20 percent of birth certificates are missing information on father's education, in part because many mothers choose not to report it. While births to unmarried parents rose consistently over the past 50 years, paternity acknowledgement initiatives helped to decouple the unmarried and missing dads trends, so that today, births with missing information on fathers are roughly 15 percent of births whereas births to unmarried parents are 40 percent of all births (Figure 2.6; Rossin-Slater (2017)). These mothers are particularly disadvantaged (Table 2.2): the mothers are younger, much less likely to be white, and only 8 percent are married. The mothers are also less educated and their babies have lower birthweight than births where father's education is reported. These births most closely resemble births to fathers who have not finished high school (Column (4)), so, to address the missing dads issue, I calculate trends where I assume all of the missing dads have less than a high school degree.

I plot these trends in Figure 2.2. When I categorize missing dads as less than high school (rather than dropping these parents from the sample), I find higher levels of assortative matching starting around 1990 (the solid blue line), which is consistent with a story of high school dropouts being particularly disadvantaged, and, moreover, that these disadvantaged individuals are increasingly isolated from those with higher levels of education.

2.1.3 Trends by race

In Figure 2.7, I plot trends in assortative matching for couples where the woman is black. In over 90 percent of these couples, her partner is also black. Sorting among

these couples has been on a slight upward trajectory over time, and trends largely agree across the samples I examine. In particular, sorting trends among married black couples are similar to sorting trends among the full sample of black couples. Trends for white couples, not very surprisingly, resemble trends for the full sample of couples (Figure 2.8). In Figure 2.9, I compare trends for white couples vs. black couples, but just for the CPS sample (to reduce clutter). Educational sorting among black couples has, for the most part, been lower than sorting for white couples.

2.1.4 Trends by birth order

Unmarried women might match with someone of lower education for their first birth but are able to find better matches as they get older. I find that, for the most part, sorting among first vs. second births have largely coincided, although, in recent years, have separated somewhat so that first births are more assortative than later births (Figure 2.10).

2.2 Conclusion

In this paper, I developed new tools to measure educational assortative matching and applied these tools to measure sorting in a new use of the Vital Statistics birth certificates data. In particular, I showed (using a simulation exercise) that existing methods that rely on comparing observed matches to random matching to control for changing educational attainment (the "random matching normalization") can control for changing *levels* of education but fail to control for a changing gender *gap* in education.

Given this finding, I proposed a new measure of assortative matching, which I called the Perfect-Random normalization. The Perfect-Random normalization continues to compare observed matches to random matching but adds in an additional

comparison to a counterfactual world where men and women match *perfectly* according to education. The changing gender gap in education limits the maximum assortativeness that a pool of couples can attain: If men and women are very mismatched in terms of education, not everyone can find a spouse of the same education level; conversely, if they have exactly the same education distribution, then everyone can find a spouse of the same education level. Thus, the *theoretical maximum* assortativeness is a function of the gender gap in educational attainment. The random matching normalization fails to consider this upper bound, and, as a result, the measured increase in assortative matching over the past 50 years was in part mechanically driven by an increasing upper bound assortativeness as women caught up to men in educational attainment.

Using the Perfect-Random normalization, I then documented that educational assortative matching has actually *declined* over time, until around 2000, at which point it began to increase. 2000 was a pivotal year in my sample because it is the year in which men and women were closest in terms of educational attainment. Thus, my finding that assortative matching declined over time was in part a result of the theoretical maximum assortativeness increasing, and the observed assortativeness not keeping apace. As women surpassed men in educational attainment in 2000, this theoretical maximum assortativeness has again been declining, thus driving the *increase* in assortative matching as measured by the Perfect-Random normalization. At the education-category level, I found that college graduates and high school dropouts are the most assortative education groups, matching at about 50 to 60 percent of the theoretical maximum since 1960.

A final contribution of this paper was to measure trends in assortative matching using information on parental education in the Vital Statistics Birth Certificates dataset. Existing work measuring assortative matching has focused on married or cohabiting couples because of data constraints: commonly used nationally represen-

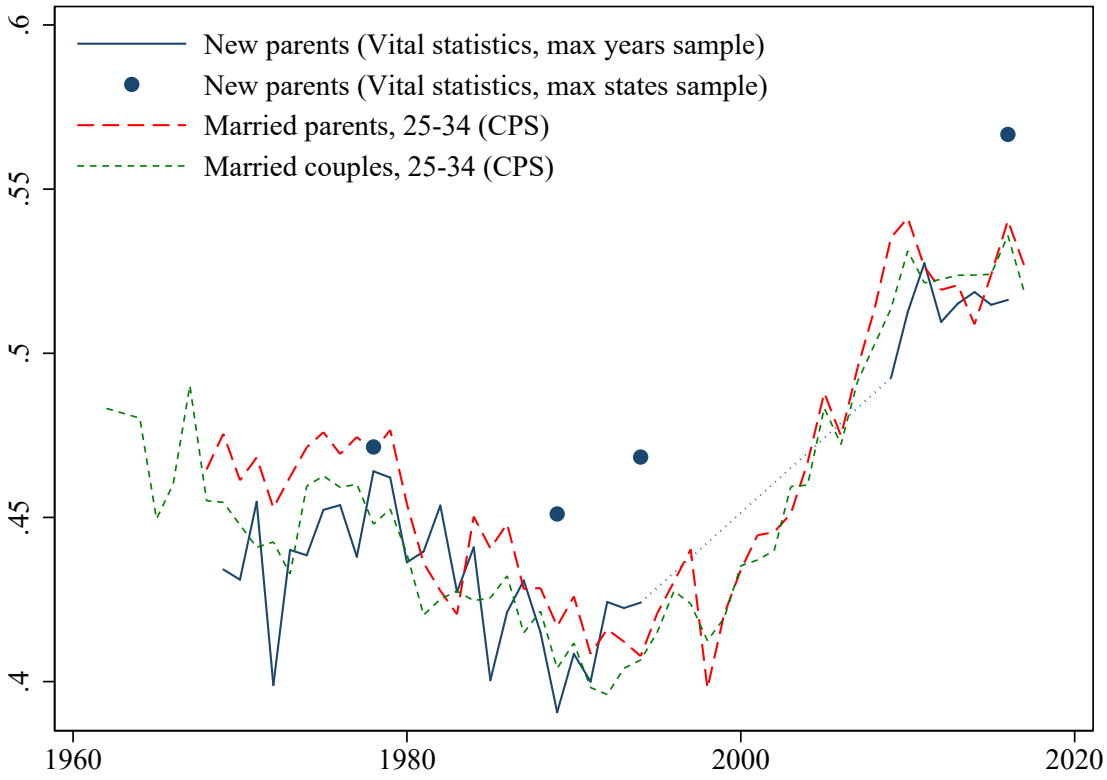
tative datasets such as the Current Population Survey or the Decennial Census are household-level datasets, and, as such, can only track couples that live together. With births to unmarried parents comprising 40 percent of all births, however, studying married couples misses a large relevant swath of the population, particularly if we are interested in studying assortative matching because of its potential implications for intergenerational mobility. I found that trends in assortative matching among all new parents in the birth certificates data were similar to married parents in the CPS, although when I take into consideration new parents where a new mother has declined to fill in information on the father (by assuming these fathers have not finished high school, as these births most closely resemble those births), I find higher levels of sorting than for married parents only.

Ample future research opportunities exist to build upon my findings in this paper. First, as I motivated in the introduction, with key progress in accurately measuring assortative matching, we can start to think about how these trends interact with dynamics of family income inequality and with intergenerational mobility. As I documented in the introduction, assortative matching is particularly associated with negative outcomes at the lower end of the education distribution: A child is *particularly* likely to not finish high school himself if both of his parents have not finished high school. Future research could explore the causal relationships behind this association. Moreover, a growing body of research has documented the existence of a gender norm that a husband should earn more than his wife (*e.g.,* Bertrand *et al.*, 2015b). As women have surpassed men in educational attainment, perhaps some couples where the woman is more educated than the man are not forming (whereas the reverse would have formed in the world where men were more educated than women). Finally, measuring educational assortative matching is very similar to measuring intergenerational educational or occupational mobility. Future research could apply the same Perfect-Random normalization to see if measured trends in intergen-

erational occupational or educational mobility are similarly driven by a changing gap in occupation or education distributions between father and son generations.

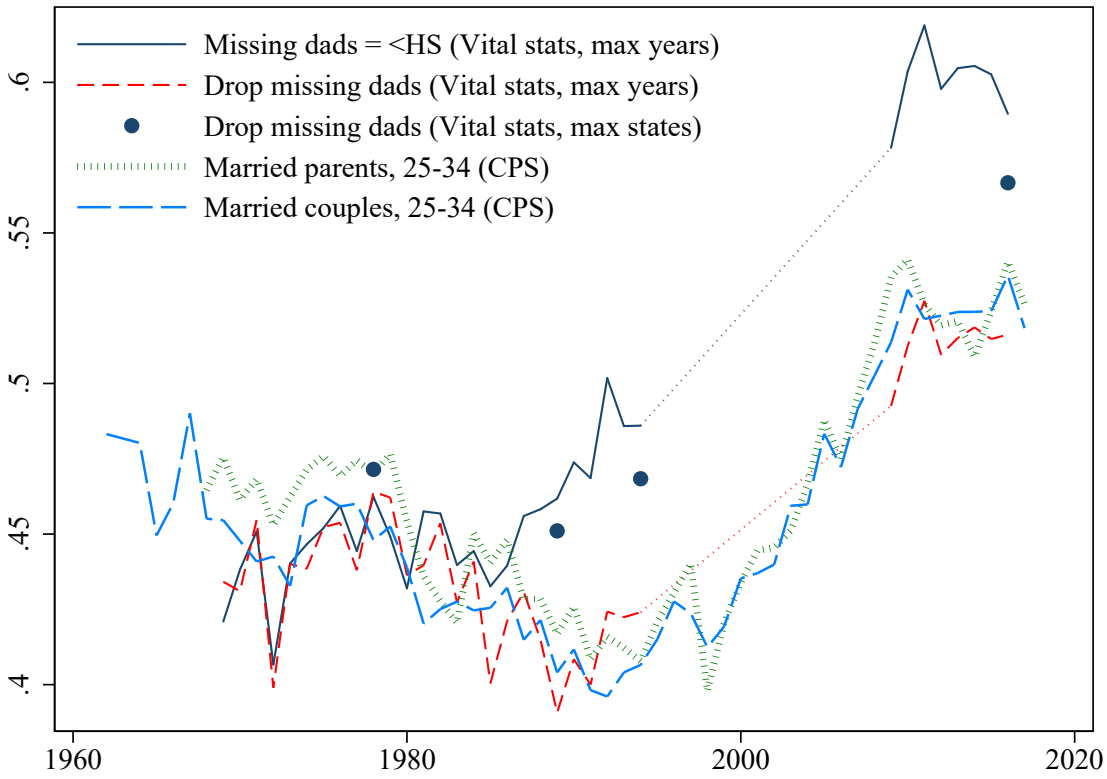
Figures

Figure 2.1: Trends in educational assortative matching, all parents vs. married parents only



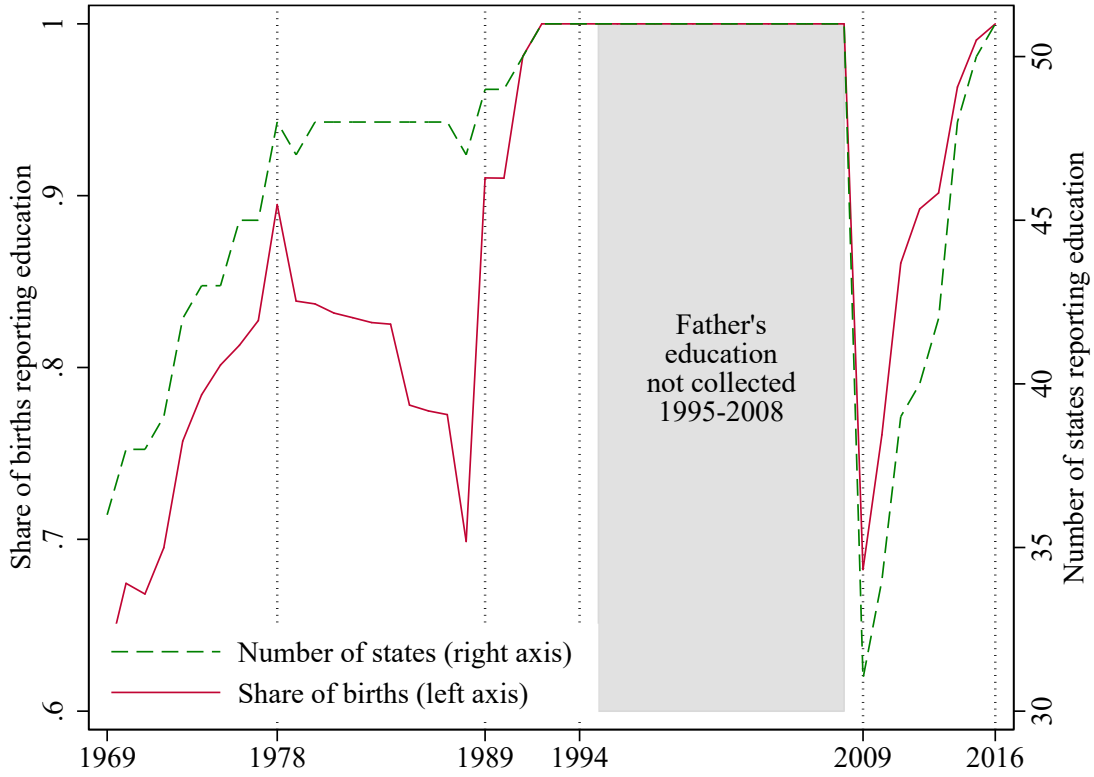
Source: Current Population Survey, ASEC/March supplements; Vital Statistics Natality
Notes: CPS sample comprises all married couples where at least one spouse is between the ages of 25-34. The CPS sample comprises all married couples where at least one spouse is between the ages of 25-34. The vital statistics max years 1% sample consists of a panel of states that consistently report parental education from 1969 to present day (but is drawn from a 1% random sample of births data from every year for computational ease). Vital statistics max states sample is a smaller set of five years (1978, 1969, 1994, 2009, 2016) where a relatively large share of states reported parental education (to maxmize geographic coverage). Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling).

Figure 2.2: Trends in educational assortative matching, missing dads = less than high school



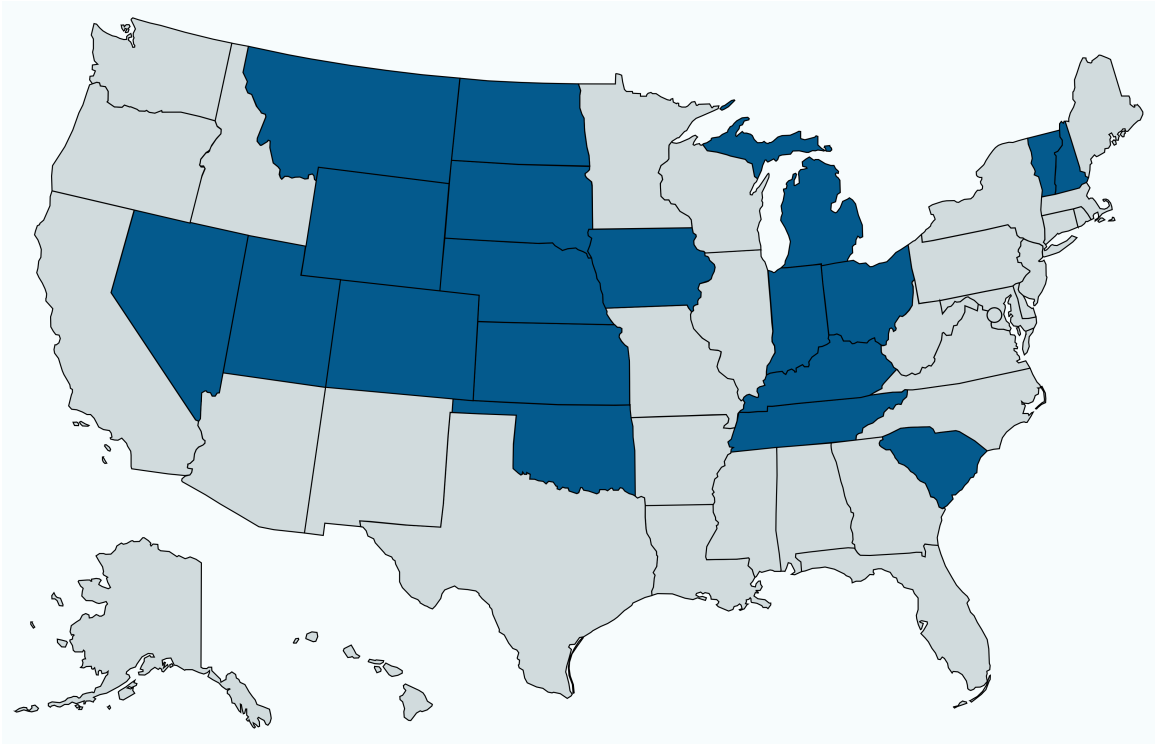
Source: Current Population Survey, ASEC/March supplements; Vital Statistics Natality
Notes: CPS sample comprises all married couples where at least one spouse is between the ages of 25-34 and at least one child under the age of 5 is in the household (to be comparable with the *flow* of new couples into forming families). The CPS sample comprises all married couples where at least one spouse is between the ages of 25-34. The vital statistics max years 1% sample consists of a panel of states that consistently report parental education from 1969 to present day (but is drawn from a 1% random sample of births data from every year for computational ease). Vital statistics max states sample is a smaller set of five years (1978, 1969, 1994, 2009, 2016) where a relatively large share of states reported parental education (to maxmize geographic coverage). Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling).

Figure 2.3: Share of births / number of states reporting parental education over time



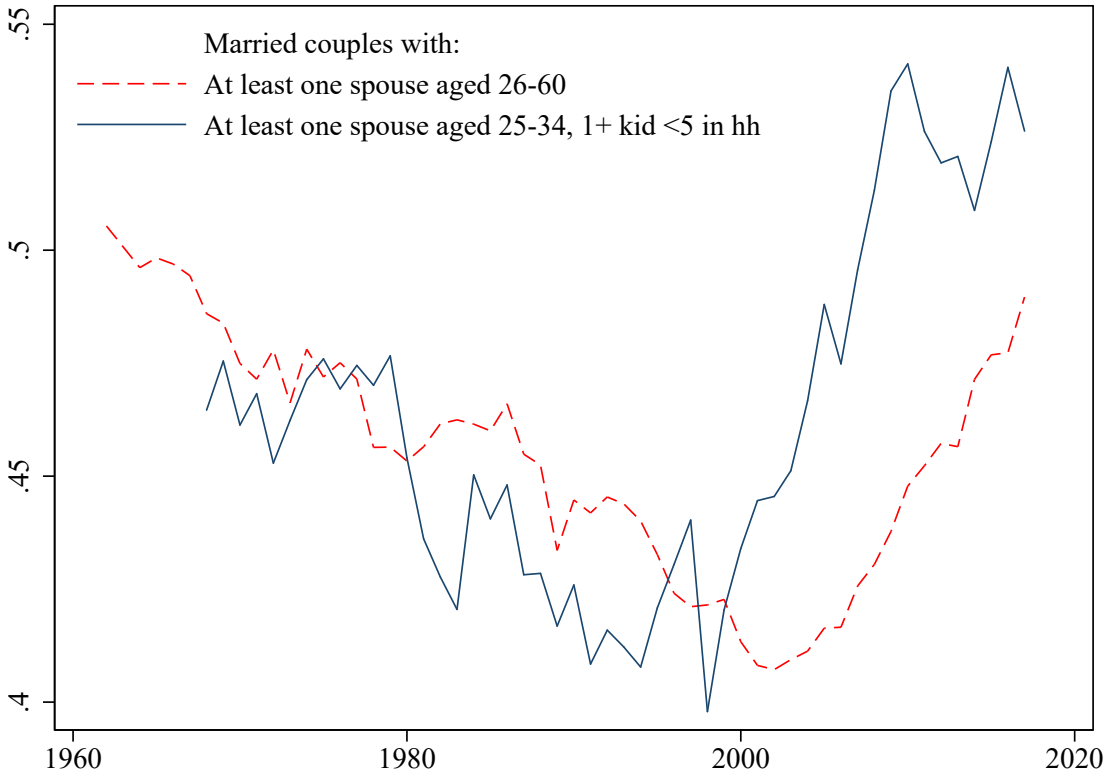
Source: Vital Statistics Natality
Notes: The vertical dotted lines indicate the years I select for the max states sample. I select them because they are years in which the share of births represented peaks. For example, in 1978, California temporarily reports parental education for one year, and then drops out of the sample in 1979, which corresponds to a spike in the share of births with parental education in 1978 in the graph above.

Figure 2.4: States in the max years sample



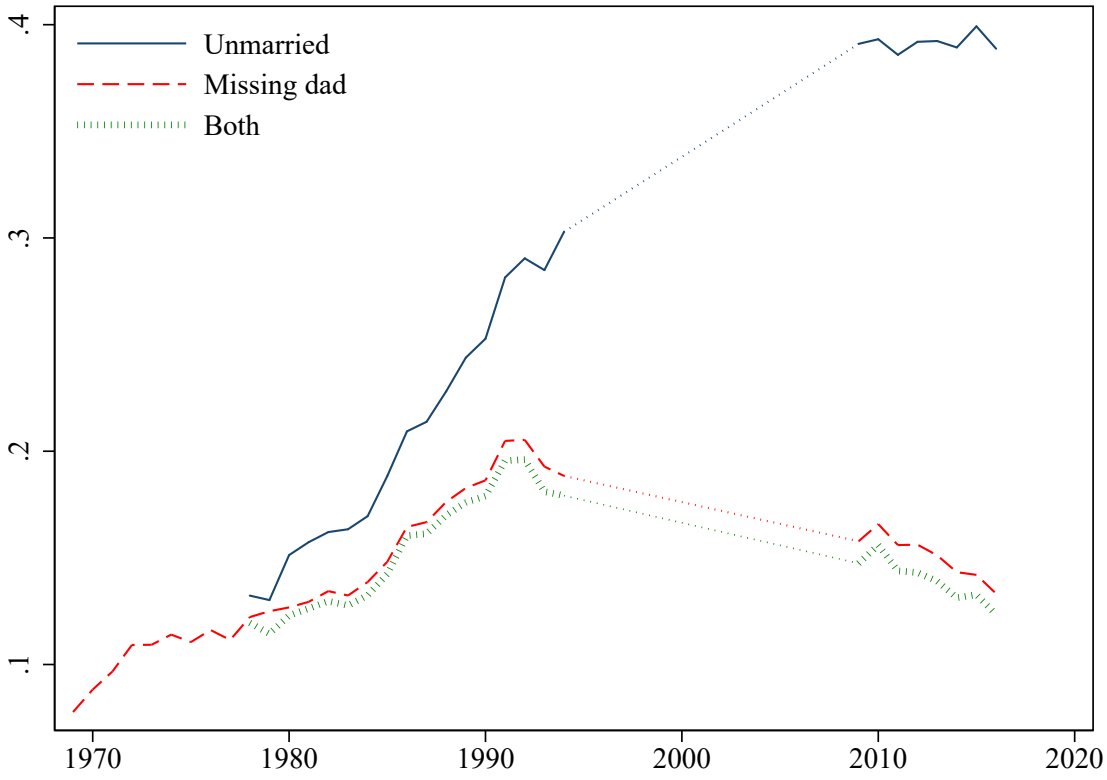
Notes: States shaded in blue consistenly report mother's and father's education between 1969 (the first year of available data) to present day. These states comprise the "max years" sample that I analyze in the Vital Statistics Natality dataset.

Figure 2.5: Trends in educational assortative matching as measured by the Perfect-Random normalization, simulation exercise vs. births samples, CPS



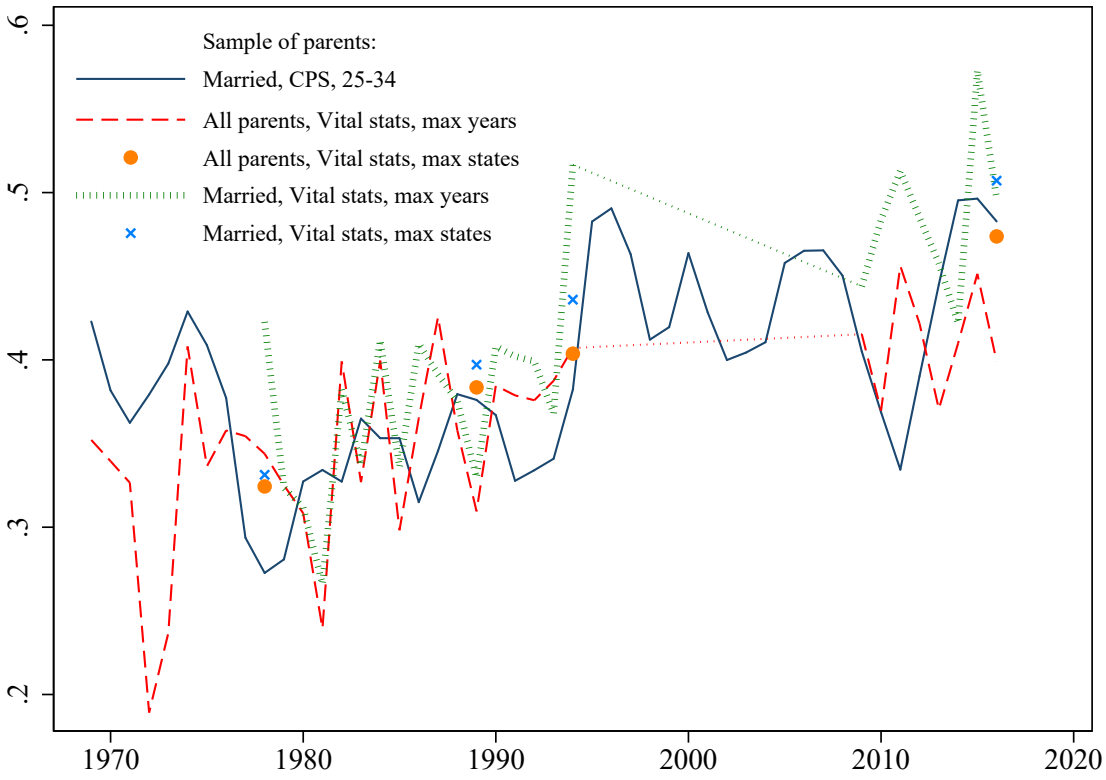
Source: Current Population Survey
Notes: Sample for dashed red line is all married couples (with spouse present) in the CPS where: (1) At least one spouse is aged 26-60 (the sample for the simulation exercise). Sample for solid blue line is all married couples (with spouse present) in the CPS where: (1) At least one spouse is aged 25-34 and at least one child under the age of 5 is present in the household, to compare trends with trends measured in the Vital Statistics births data. Education is measured according to: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling).

Figure 2.6: Share of births where mother is unmarried and/or where father's education is missing



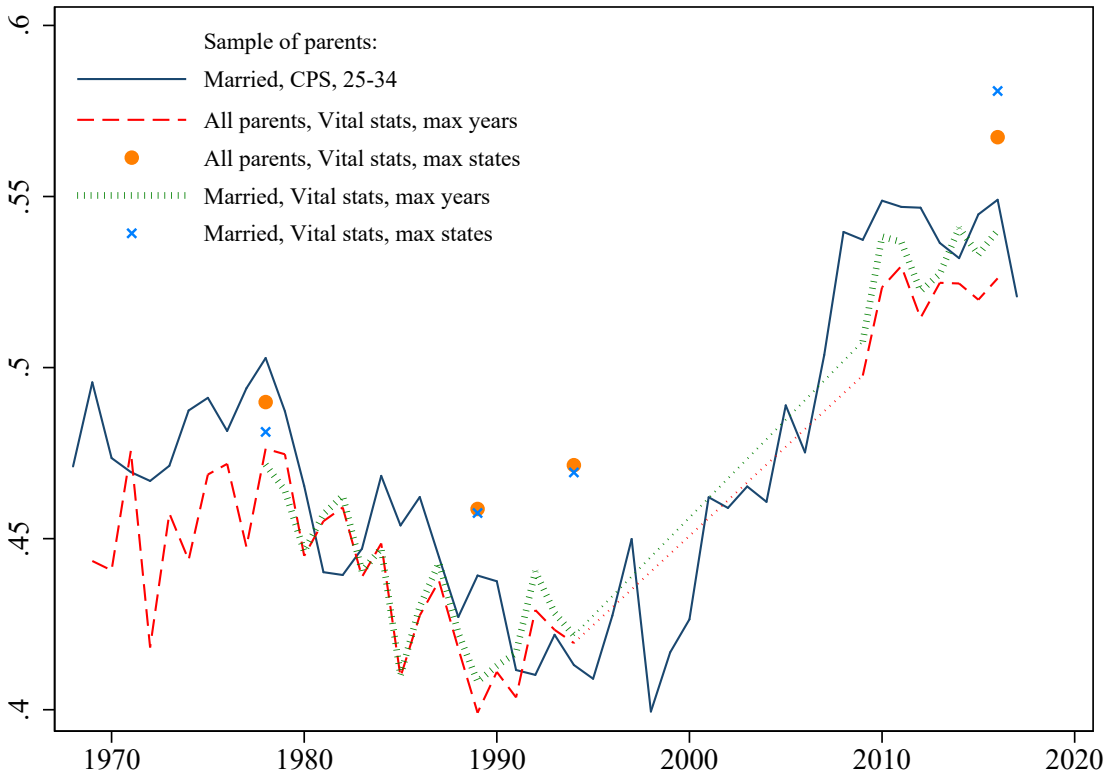
Source: Vital Statistics Natality, Max years 1% sample
Notes: Missing dad is defined as births where mother's education is reported but father's education is not. Sample is all births in states in the max years sample.

Figure 2.7: Trends in educational assortative matching for couples where the woman is black



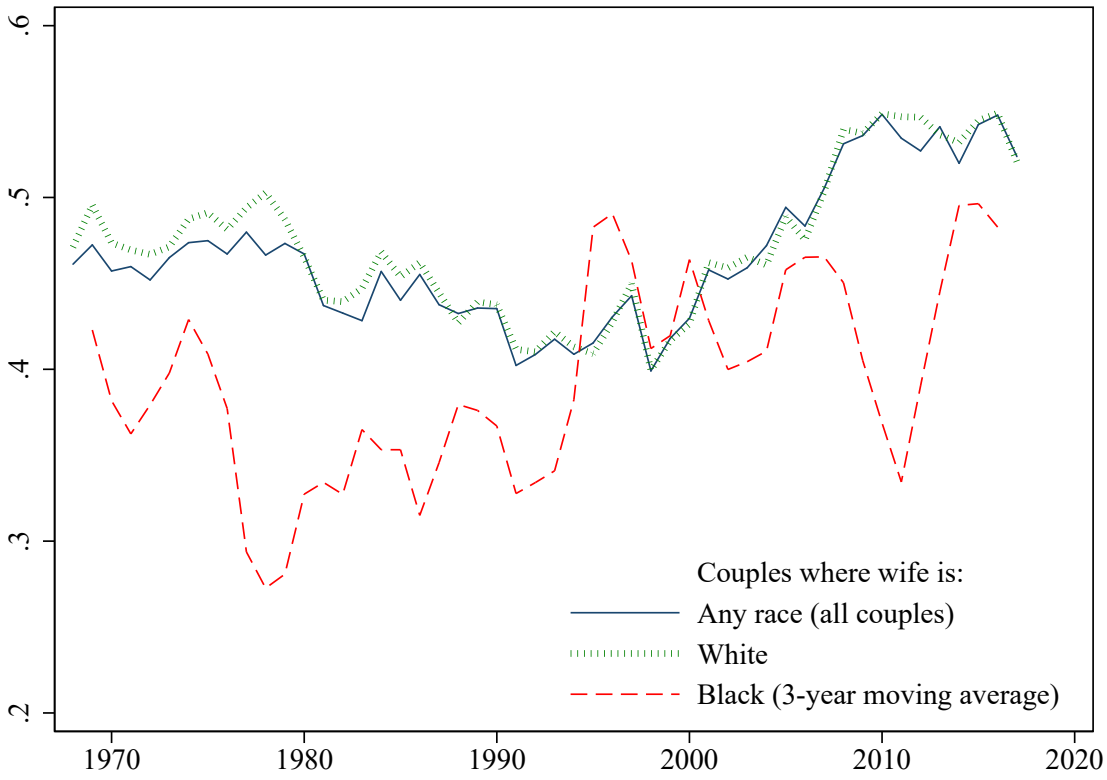
Source: Current Population Survey, Vital Statistics
Notes: For all samples, I restrict to couplles where the female partner is black. The CPS sample comprises all married couples where at least one spouse is between the ages of 25-34, at least one child under age 5 is present in the household. The vital statistics max years 1% sample consists of a panel of states that consistently report parental education from 1969 to present day (but is drawn from a 1% random sample of births data from every year for computational ease). Vital statistics max states sample is a smaller set of five years (1978, 1969, 1994, 2009, 2016) where a relatively large share of states reported parental education (to maxmize geographic coverage). All births in the given years are included in the max states sample. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling). Trends for black couples in the CPS are smoothed according to a 3-year moving average because they are a relatively small sample in the CPS and thus are noisy.

Figure 2.8: Trends in educational assortative matching for couples where the woman is white



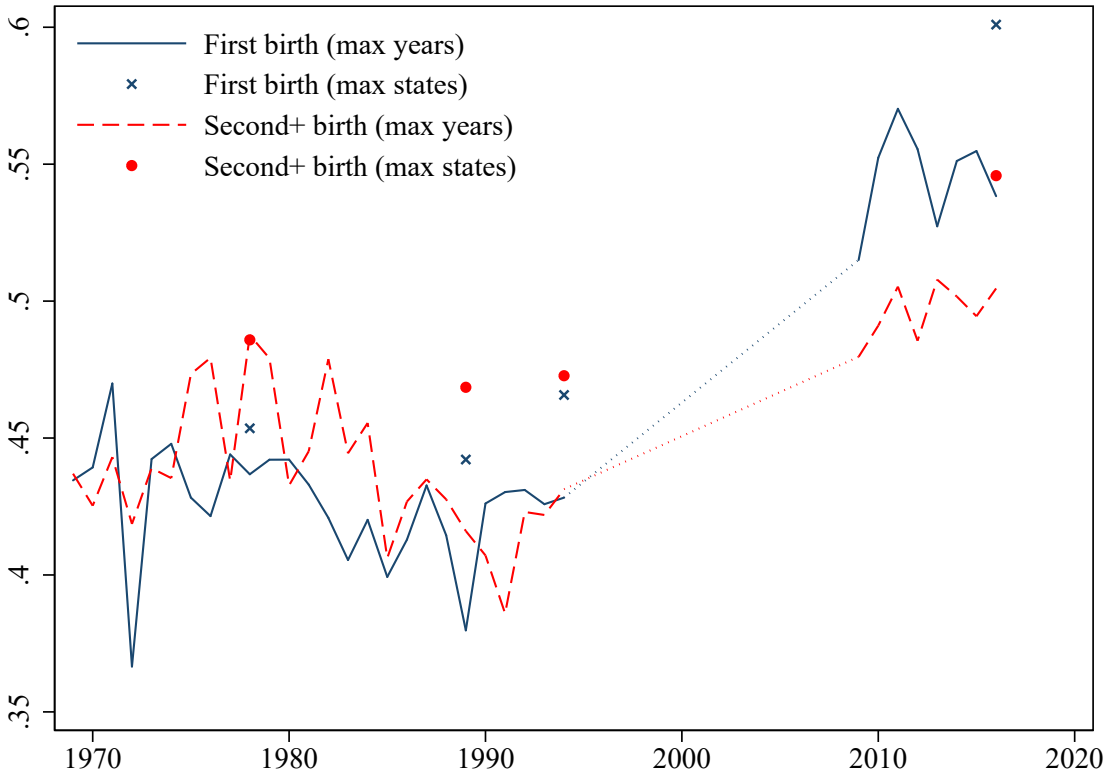
Source: Current Population Survey, Vital Statistics
Notes: For all samples, I restrict to couplles where the female partner is white. The CPS sample comprises all married couples where at least one spouse is between the ages of 25-34, at least one child under age 5 is present in the household. The vital statistics max years sample consists of a panel of states that consistently report parental education from 1969 to present day (but is drawn from a 1% random sample of births data from every year for computational ease). Vital statistics max states sample is a smaller set of five years (1978, 1969, 1994, 2009, 2016) where a relatively large share of states reported parental education (to maxmize geographic coverage). All births in the given years are included in the max states sample. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling).

Figure 2.9: Trends in educational assortative matching by race of wife



Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 25-34. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some college education (13-15 years of schooling); 4 = College graduate (16+ years of schooling). Trends for black couples in the CPS are smoothed according to a 3-year moving average because they are a relatively small sample in the CPS and thus are noisy.

Figure 2.10: Trends in educational assortative matching by whether first or later birth



Source: Vital Statistics Natality
Notes: The vital statistics max years 1% sample consists of a panel of states that consistently report parental education from 1969 to present day. Vital statistics max states sample is a smaller set of five years (1978, 1969, 1994, 2009, 2016) where a relatively large share of states reported parental education (to maxmize geographic coverage). All births in the given years are included in the max states sample.

Tables

Table 2.1: Summary statistics, by sample

		Vital Statistics Births		
	(1) CPS	(2) 1% sample	(3) Max years (1%)	(4) Max states
Father's age	31.61	29.65	28.85	29.89
Mother's age	28.95	26.55	25.80	26.79
Mother white	0.87	0.79	0.84	0.78
Mother black	0.07	0.16	0.13	0.16
Father white	0.87	0.70	0.75	0.70
Father black	0.08	0.11	0.08	0.11
Mother married	1.00	0.68	0.73	0.69
–Mother's education				
Less than high school	0.14	0.22	0.21	0.21
High school graduate	0.38	0.35	0.39	0.35
Some college	0.24	0.22	0.23	0.23
College graduate	0.24	0.21	0.18	0.22
–Father's education				
Less than high school	0.15	0.18	0.16	0.17
High school graduate	0.35	0.37	0.39	0.36
Some college	0.23	0.21	0.22	0.21
College graduate	0.27	0.25	0.23	0.26
Observations	255,832	1,581,042	274,621	14,119,495

Source: Vital Statistics Natality, CPS
Notes: The CPS sample comprises all married couples where at least one spouse is between the ages of 25-34 and at least one child under age 5 is present in the household. I analyze a 1% random sample of births for computational ease (column (2)). The vital statistics max years 1% sample consists of a panel of states that consistently report parental education from 1969 to present day. Vital statistics max states sample is a smaller set of five years (1978, 1969, 1994, 2009, 2016) where a relatively large share of states reported parental education (to maxmize geographic coverage). All births in the given years are included in the max states sample.

Table 2.2: Summary statistics of mothers and children by father's education

	Father's education				
	(1) Coll+	(2) SC	(3) HS	(4) < HS	(5) Missing
Mother's age	29.52	26.99	25.10	24.17	22.65
Mother white	0.93	0.89	0.88	0.87	0.57
Mother black	0.04	0.08	0.10	0.10	0.39
Mother married	0.97	0.87	0.82	0.73	0.08
–Mother's education					
Less than high school	0.01	0.05	0.17	0.56	0.44
High school graduate	0.13	0.32	0.59	0.35	0.38
Some college	0.26	0.44	0.18	0.08	0.15
College graduate	0.60	0.19	0.06	0.01	0.02
–Child birthweight					
Birthweight (grams)	3403.98	3369.27	3337.72	3264.84	3143.18
Low birthweight (< 2500g)	0.05	0.05	0.06	0.09	0.12
Very low birthweight (< 1500g)	0.01	0.01	0.01	0.01	0.02
Observations	53,868	50,048	90,358	37,215	43,132

Source: Vital Statistics Natality
Notes: Sample is the Vital Statistics 1% max years sample.

Chapter 3

The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?

Co-authored with Ilyana Kuziemko, Jessica Pan, and Ebonya Washington[1]

3.1 Introduction

The entry of women, especially mothers, into the labor force in high-income countries has been one of the most transformative economic and social developments of the past century. However, in the US, this trend has plateaued since the 1990s (see Figure 3.1), and even in countries where it has continued to increase, it has been driven by part-time work (Blau and Kahn, 2013). This "stalling" of male-female convergence in the labor market occurs *despite* the fact that over this period women have increased their

[1]We thank seminar participants at Aarhus University, Cornell, Dartmouth, GRIPS, NBER Summer Institute (joint Labor Studies-Childrens meeting), NUS, Princeton, SOLE, Tinbergen Institute, UCLA Anderson, Yale-NUS. We benefited from helpful discussions with Claudia Goldin, Raquel Fernandez, Henrik Kleven, Suresh Naidu, Emily Oster, and Richard Rogerson. Courtney Burke, Elisa Jacome, Rais Kamis, Tammy Tseng and especially Elena Marchetti-Bowick have provided excellent research assistance.

human capital in the form of education (they are now significantly more educated than men, see Figure 3.2) and in the form of job experience by delaying childbirth (Appendix Figure A.1).

In this paper, we propose a novel resolution to this puzzle. We hypothesize that modern cohorts of women *underestimate the employment costs of motherhood*, which we loosely define as the time, effort or money required for them to raise their children in a manner they deem appropriate, while also working outside the home. Examples of these costs might include the emotional cost of being separated from the child while at work, the per-hour cost of a nanny or day-care service, guilt over (perceived or real) underperformance as an employee or mother, or diminished sleep or other aspects of wellbeing due to working while also performing child-care activities.[2] If women have assumed that these costs have fallen when in fact they remain high, then women would acquire costly human capital (because they assume they will enjoy the returns over a long period of labor-force attachment), but also exhibit diminished labor supply after motherhood (because, in fact, the employment costs of motherhood remain high, leading to withdrawal from the labor market). We build support for this hypothesis in several steps, using a variety of datasets from the US and the UK.

We begin by documenting in an event-study framework the evolution of employment outcomes before and after the "event" of the birth of a first child. Women's probability of employment, relative to a no-baby counterfactual, declines by roughly thirty to forty percentage points upon motherhood in the US and UK, effects that are substantially larger than the effects recently found in Scandinavian countries (Kleven *et al.*, 2018; Angelov *et al.*, 2016). This decline is essentially fully realized the year of the first child's birth, and exhibits very little recovery in the five- to ten-year post-period we can observe. Heterogeneity in the employment response to children is also

[2]See Bertrand (2013), who finds that among college-educated mothers, those with a career report being more unhappy, stressed, and tired than stay-at-home mothers. See Fortin (2005) for a longer discussion of feelings of guilt among working mothers.

quite similar in the US and UK. For example, a college education has a modest "protective" effect on employment, but even these more educated women exhibit large and significant "mommy effects" of decreased employment.

We argue that women *underestimate* these large employment effects, in both the short- and longer-run, and the bulk of the empirical work presents evidence supporting this claim. If the effect of motherhood on the ability to maintain both work and family commitments is indeed unexpected in the short-run, then women's *beliefs* about the appropriate balance between home and market work should change discontinuously upon the birth of the first child, as this event allows them to update their beliefs with new information. We thus examine how attitudes on gender roles and work-family balance (e.g., agreement with statements such as "a woman and her family would all be happier if she goes out to work" or "both husbands and wives should contribute financially") change once a woman has a child, using an event-study framework. Whereas before motherhood most women say that work does not inhibit women's ability to be good wives and mothers, after the birth of their first child they become significantly more negative toward female employment, consistent with having underestimated the impact of motherhood on employment. The evolution of gender-role attitudes also roughly approximates a step-function: no meaningful pre-trends, a steep change around the time of the first child's birth, and no recovery in the medium-run. Interestingly, the heterogeneity results are quite different than those for employment: while educated mothers experience smaller employment declines than their less educated counterparts, they exhibit larger updates in their beliefs (in the direction that working women hurt family life).

We interpret the event-study results for the work-family attitudes as suggesting that women have underestimated the employment effects of motherhood—as upon motherhood they become more pessimistic about women's ability to both work outside the home and fulfill their perceived responsibilities at home—and that this ten-

dency is especially pronounced for educated women. We find additional evidence for this claim from surveys that ask parents retrospective questions. The large majority of women agree that "being a parent is harder than I thought it would be," and this opinion is concentrated among college-educated women. For men, a far lower share agree with that sentiment, and there is no difference by education.

The evidence described so far suggests that women underestimate the employment effects of motherhood, but for this underestimation to help explain the rise in female human capital accumulation (as we claim it does), then they would have to underestimate these effects while they are making their key education decisions. We thus examine the expectations of young women around the age of eighteen. Goldin (1990) and Goldin (2006b) had earlier demonstrated that teenage women in the US substantially *underestimated* their future labor supply, when asked in 1968 about their likelihood of employment in their thirties. We show that this result has now completely reversed for more recent cohorts. Even by 1978, only ten percent of female high school seniors said they would be home-makers at age thirty, a dramatic decline over a ten-year period and a roughly ten-point underestimate relative to their *actual* experience twelve years later. From roughly 1985 through the present, no more than two percent of female high school seniors say they expect to be home-makers at age thirty, even though the share of home-makers among thirty-year old women has remained roughly constant at 15-18 percent since 1990. Importantly, these same young women expect to be mothers (of two or three children), but not, apparently, stay-at-home mothers.

That recent cohorts of women seem to systematically underestimate the employment consequences of motherhood—and thus overestimate their future labor market attachment—begs the question of how they could get such an important prediction so substantially wrong. After all, motherhood is a very common event and they could in principle learn from their own mothers' experience and that of peers. We close

the paper with our preferred explanation: *that the employment costs of motherhood have risen in an unexpected manner.* While we do not formally test and reject other explanations, we provide a variety of evidence that supports this proposed increase in costs.

We start by presenting a simple model of women's education and employment choices when the future employment costs of motherhood are uncertain. Women form predictions over the level of their own future costs by observing their own mothers, but they inherit the costs their mothers faced with some noise (which may or may not have a mean of zero). We first show that the model yields two of our most striking empirical results: that educated women exhibit smaller employment "mommy effects" than their less educated counterparts, but at the same time appear the most "surprised" by the high employment cost of motherhood (in terms of their greater updating of beliefs about work-family gender roles and their larger retrospective expressions of surprise at how hard parenthood is). Women in our model choose higher levels of education—which raises the return to market work *ex post* and thus results in higher labor supply post-motherhood—in part *because* they underestimate the cost of motherhood *ex ante*. We show that these results hold *regardless* of whether, on average, the employment costs of motherhood have increased or decreased for the current generation relative to the previous one. That these results hold under relatively general conditions not only bolsters the credibility of the model, but also serves as a "proof of concept" for the empirical results. We then show that some of our other central results—that *unconditional on education* women in modern cohorts appear to underestimate the costs of motherhood—holds if and only if costs have on average *increased* for today's mothers relative to earlier cohorts.

To complement this theory-based argument that the cost of motherhood must have risen, we provide a collage of empirical evidence suggesting that, while the cost likely fell during much of the twentieth century, it appears to have recently risen along some

important metrics, especially since the 1990s. Some of this evidence we take from past literature and some, to the best of our knowledge, we provide for the first time. For example, past work has emphasized the *decline* in the cost of motherhood over the middle decades of the 20th century due to technological advances (e.g., household appliances and advances in infant formula) and links these falling costs to the large rise in female LFPR over the same period.[3] We argue that those advances may have fully played out and in fact more recent developments (e.g., research advocating the benefits of breastmilk over formula) may have effectively increased the costs of motherhood. Indeed, in both the US and the UK, we show that breastfeeding rates have increased significantly over the past several decades, especially since the 1990s. Similarly, while the underlying cause is still debated, the time cost associated with child care beyond infancy has been rising, especially for educated mothers (see Ramey and Ramey, 2010 and Guryan *et al.*, 2008 for evidence from the US and Borra and Sevilla, 2015 for the UK). Consistent with these trends, we show that large majorities of mothers tell survey takers that motherhood is harder today than for their own mothers and that they are more involved in their children's lives than their mothers were in theirs. In the conclusion, we briefly take up the question of *why* the cost of motherhood may have increased over recent decades.

Our paper contributes to the literature on the employment effects of motherhood, a topic of long interest among labor economists. Perhaps surprisingly, the number of papers that have performed parallel analysis using the birth of a first child in an event-study analysis of labor supply measures is quite limited.[4] Recent event-study

[3]We review this literature in more detail in Section 3.7.

[4]Papers that have used other methods to examine the question of children and women's labor supply include, of course, Angrist and Evans (1998), who focus on developing credible instruments for the birth of a *third* child. More recently, Lundborg *et al.* (2017) argue that, conditional on undergoing *in vitro* fertility (IVF) treatment, its success is sufficiently random to serve as a valid instrument for fertility. While event-studies cannot formally address endogeneity of fertility, interestingly, Kleven *et al.* (2018) shows that the event-study estimate of the effect of a *third* birth lines up closely to the IV-based estimate using twins or sex composition as an IV for a third birth. A small set of papers have examined the effect of childbirth using a woman fixed-effects strategy but have not traced out

analyses by Kleven *et al.* (2018) and Angelov *et al.* (2016) using rich administrative data from Denmark and Sweden, respectively, uncover persistent declines in women's relative earnings and labor supply at the onset of parenthood. Similar studies for the US include Byker (2016), who uses the SIPP to examine how motherhood impacts women's labor supply in the short run, Chung *et al.* (2017), who link SIPP data SSA earnings records to examine the impact of parenthood on the earnings gap between mothers and fathers, and Goldin and Mitchell (2017), who use a variety of data sources to show employment effects of motherhood across cohorts. Bertrand *et al.* (2010) focus on female MBA graduates and show that these highly educated women reduce their labor supply in the year of first birth, with further reductions taking place over the next four years.

We also contribute to a large economics literature on gender-role norms, and in particular how they change over time.[5] For example, Goldin (2006b) suggests that innovations in contraception may have contributed to altering women's identity in the 1960s and 1970s. Fernández *et al.* (2004) argue that men growing up in families with working mothers appear to have developed more liberal gender-role attitudes.[6] Related, perhaps, to the slowdown in labor market convergence we noted earlier, Fortin (2005) shows that while birth cohorts have become more liberal regarding norms about women in the workplace, this trend has recently plateaued in OECD countries. We show that gender-role attitudes appear to exhibit life-cycle effects: women themselves adopt more traditional attitudes after they become mothers.

---

the effects in event-time, so have less ability to untangle coincident trends in labor supply and family planning (Waldfogel, 1997; Wilde *et al.*, 2010).

[5]Recent contributions to this literature have focused on norms in the marriage market or within couples. Bertrand *et al.* (2015a) and Kleven *et al.* (2016) provide a variety of evidence that couples follow the norm that husbands should out-earn their wives. Bursztyn *et al.* (2017) show that traditional gender-role norms remain relevant even among career-oriented MBA students—anticipation of the marriage market appears to limit the public signaling of career ambitions of single female MBA students.

[6]On the other hand, Alesina *et al.* (2013) highlight the long-term persistence of gender norms, showing that ethnicities and countries whose ancestors practiced plough cultivation (which is more suited to male labor) have lower rates of female labor force participation even today.

Of course, the idea that motherhood changes women's priorities is hardly original, and appears frequently in the popular press.[7] To the best of our knowledge, we are the first to examine how gender-role norms regarding employment change after parenthood in a formal event-study framework. We are aware of a small number of academic papers, mostly from sociology, that examine similar questions, all using panel data from high-income countries (Grinza *et al.*, 2017; Baxter *et al.*, 2015; Shafer and Malhotra, 2011; Jarrallah *et al.*, 2016; Vespa, 2009). But these panels tend to be short (in some cases, only two waves), and thus these analyses cannot perform event studies to determine if any estimated effect on attitudes is coincident with motherhood or instead driven by trends.

The remainder of the paper is organized as follows. Section 3.2 introduces our main data sources. Section 3.3 describes the empirical strategy. Section 3.4 reports our results on the effect of parenthood on labor supply. Section 3.5 provides a variety of evidence that, in the short and medium run, women underestimate the employment effects of motherhood. Section 3.6 examines the accuracy of their longer-run expectations. Section 3.7 relies on both theoretical and empirical evidence to argue that the reason why modern women have not anticipated the costs of motherhood is because these costs recently increased in an unpredictable manner. Section 3.8 concludes and poses some questions for future work.

3.2 Data sources

In this section, we describe our main data sources and our sampling restrictions. A substantial portion of the empirical work involves event-study analysis. We thus adopt sample restrictions with that exercise in mind. Our preferred sample will include only those individuals whom we observe at least once before and once after the "event"

[7] See, e.g., "What Happens to Women's Ambitions in the Years After College," in *The Atlantic*, December 2016. `https://www.theatlantic.com/business/archive/2016/12/ambition-interview/486479/`.

of the birth of a first child, so that all subjects help identify a standard difference-in-difference estimate of the effect of parenthood (though in robustness checks we will add back in subjects who remain childless or had children before entering the sample, to serve as controls). We also, whenever possible, require that we observe respondents over the full period of their most likely child-bearing years (between ages 20 and 40).[8] Again, in robustness checks, we relax this restriction.

Below, we briefly describe our four main panel data sources. In the interest of space, we relegate greater detail on variable construction, weighting and sampling restrictions to Appendix A.2.

3.2.1 British Household Panel Survey

The British Household Panel Survey (BHPS) is a longitudinal survey that runs from 1991 through 2009. Col. (1) of Table 3.1 shows summary statistics for women in our preferred sample in the BHPS. The sample restrictions described above leave us with a total of 681 women. On average, in our analysis sample, we observe individuals 4.1 times pre-birth and 6.6 times after. While we do not impose perfect balance for our event-time figures, it is reassuring to know that we observe parents several times before and after their first birth so that trends are unlikely to be explained by composition changes, though we return to this concern in Section 3.4.2. While not our focus, Appendix Table A.1 provides summary statistics for the men in the BHPS as well for the men in our U.S. panel data sources.

[8]The one exception is the British Household Panel Survey (BHPS). As the BHPS only spans 18 years (with a later additional booster sample spanning only about 10 years), imposing the 20-40 age restriction is impossible. Instead of imposing the 20-40 restriction, we require our main sample of women to be at least 16 years old at the start of our panel. For our control samples of childless women, we require that we observe them at some point at age 29, the median age at first birth among women in our treatment sample.

3.2.2 U.S. panel data sources

We make use of three, commonly used sources of U.S. panel data: the National Longitudinal Surveys of Young and Mature Women (hereafter NLSW68), the National Longitudinal Survey of Youth (NLSY79), and the Panel Study of Income Dynamics (PSID).

We make use only of the "young women" component of the NLSW68.[9] Women were 14-24 when first interviewed in 1968 and interviews continue through 2003. For the most part, we can only examine employment outcomes with these data, as there are no questions on gender-role attitudes that are asked in consistently over time. However, we will make use of a variable that asks women in 1968 about their employment and parenthood expectations when they are 35. Our preferred sample includes only those women whom we observe at least once before and after the birth of their first child and who remain in the sample between ages twenty and forty. This sample includes only women who were born between 1948 and 1954 (so they are roughly in their sixties today). Our preferred sampling criteria leave us with 1,413 women, though, as noted above, we will drop some of our preferred sample restrictions in robustness checks, enlarging the sample.

Summary statistics are presented in col. (2) of Table 3.1 for our preferred NLSW68 sample. Given the longer time period covered by the NLSW68 (and, in fact, all our U.S. panel data sources), we have many more observations pre- and post-baby than in the BHPS.

The NLSY79 begins interviewing young men and women between the ages of 14 and 22 in 1979 and continues through 2014. While we occasionally will show results for men, we focus on the female half of the sample. Again, to ensure that we observe the full period of respondents' most likely child-bearing ages, we only include the

[9] While the National Longitudinal Surveys of Young and Mature Women include an older ("mature women") sample, they have for the most part had their first child before entering the sample.

1959-1964 cohorts (so our sample is born roughly ten years after their NLSW68 counterparts). These sampling criteria leave us with 2,256 women. Summary statistics for this preferred NLSY79 sample appear in col. (3) of Table 3.1.

Our final U.S. panel data source is the PSID. While it begins in 1968 (and continues through today), our preferred analysis sample includes all individuals for whom we observe the first birth *after 1979* (because the employment outcomes we most use are not coded consistently before this date). To meet our usual requirement that we observe individuals between the ages of twenty and forty, our preferred sample includes only the 1957-1975 cohorts, leaving us with 1,007 women. Col. (4) of Table 3.1 displays summary statistics for our PSID sample. On average, the women in our sample are born about five (fifteen) years after those in the NLSY79 (NLSW68).

Due to slightly different definitions, we cannot compare all variables across the BHPS and our U.S. datasets, but where we can, the differences are as expected. Note that even the PSID, our most recent U.S. panel dataset, samples women who are from slightly earlier cohorts than the BHPS. Consistent with having later cohorts in the BHPS and the general tendency of the British to delay childbirth relative to Americans (recall Appendix Figure A.1), individuals in all our U.S. samples have their first child at a younger age than do women in the BHPS and also have more children in the period in which we observe them.[10] Consistent with the evidence in Figure 3.2, the summary statistics show that American women are more likely to have completed university education than their British counterparts, despite coming from earlier cohorts. American women in the PSID and NLSY79 samples are also significantly more likely to have obtained a college degree than their counterparts in the NLSW68, consistent with the rapid growth in female college graduation rates in the US between the 1940s and 1970s cohorts (Goldin *et al.*, 2006b).

[10]Our British panel is also shorter than our American panels, so the U.K. women are also less likely to have completed their fertility by the end of the panel.

It is worth noting that while these datasets represent the most recent panel data for which we can observe women over their main reproductive years (NLSY97 respondents, e.g., are still in their thirties), even the most recent birth cohorts in our analysis (those in the BHPS) are today in their forties. While we will often describe results in the present tense, it is unclear they will reflect the behavior of women who are today in their twenties and thirties.

3.3 Empirical strategy

Much of our analysis examining the short- and medium-run changes associated with parenthood makes use of a basic event-study methodology, defining the first child's year of birth as the "event." Specifically, we model a given outcome y_{it} (e.g., current employment, attitudes toward gender roles) for person i in year t as:

$$y_{it} = \sum_{\tau=-5}^{\tau=\tau^{max}} \beta_\tau \cdot \mathbb{1}[\tau = t - e^i] + \sum_a \gamma_a \cdot \mathbb{1}[a = \text{Age}_{it}] + \delta_t + \alpha_i + \epsilon_{it}. \tag{3.1}$$

We index event time (time relative to birth of a child) by τ. The variable $\mathbb{1}[\tau = t - e^i]$ is defined as follows: e^i denotes the calendar year in which person i had their first child, so $\mathbb{1}[\tau = t - e^i]$ is an indicator for person i in year t having had their first child τ years ago. Negative values of τ indicate having a first child $|\tau|$ years in the future. In the summation term, we omit the year before the first birth, and the $\tau = -5$ (or $\tau = \tau^{max}$) term in fact includes all years greater than or equal to five years before (or τ^{max} years after) the first birth. We set τ^{max} to five or ten, depending on the length of the sample period in a given panel dataset. This specification normalizes the year $\tau = -1$ to zero, though in our figures we will often add to all coefficients the raw mean of the outcome variable at $\tau = -1$ to facilitate comparisons across different subgroups in both levels and changes. We control for a vector of individual fixed

effects (α_i), a vector of calendar-year fixed effects (δ_t) and a vector of age-in-years fixed effects ($\sum_a \mathbb{1}[a = \text{Age}_{it}]$). The error term is ϵ_{it}. We cluster standard errors at the individual level.

We present our results by plotting the β_τ coefficients, to show the within-person evolution of our outcome variables relative to the event of parenthood, conditional on year and age fixed effects. We also show difference-in-difference results, which we estimate using the same specification as in equation (3.1), but with the event-time year dummies replaced by a dummy variable, $\textit{Post-Baby}_{it}$ (indicating that in year t we observe person i *after* the birth of her first child).

3.4 How does motherhood affect employment?

We begin by showing how labor-supply measures evolve before and after the event of parenthood, using both the BHPS and our U.S. panel data sources. For the main empirical results in the paper, we focus on event-study figures, as they provide more transparent and visual evidence on how outcomes evolve before and after parenthood (e.g., pre-trends, persistence) than do the difference-in-difference results. However, performing a comprehensive battery of robustness checks on each event-study graph would require exorbitant space, so in the Appendix we provide the difference-in-difference results corresponding to the event-study graphs and subject them to a battery of checks. In Appendix tables we also show how outcomes evolve after the birth of a *second* child (and whether we can distinguish this effect from that of the first child) as well as how *men's* outcomes react to the birth of a first child. In the interest of space, we only discuss some of these results, but interested readers can refer to the Appendix for a fuller treatment.

3.4.1 Main results

We begin our analysis by examining the share of women who are currently employed, before and after the birth of a first child. Figure 3.3 shows the event-study coefficients from equation 3.1, estimated on our U.K. data, which defines currently employed as having been at work last week. The year before the birth of a child, 87 percent of women are working (recall that the coefficient for $\tau = -1$ is merely the raw mean at $\tau = -1$), a rate only three percentage points lower than for men who will become fathers the next year (not shown). The year of the birth itself exhibits about a forty percentage-point decline in employment. Approximately five to ten percentage points of this decline are recovered in the following year, though no further recovery can be detected beyond that. When we instead collapse event time into a simple before and after the birth of the child in a difference-in-difference estimation (the estimate reported at the top of the figure), we find that the probability of being employed the week of the survey declines by 38 percentage points.

We now replicate our U.K. employment results using our U.S. panel data sources. For each dataset, we use an extensive margin measure of labor supply that matches as closely as possible that in the BHPS, but see Appendix A.2 for exact variable definitions. Figure 3.4 plots the event-study coefficients separately estimated in the NLSW68, the NLSY79 and the PSID. In their overall shape, the three event studies in Figure 3.4 are very similar to each other and to that from the BHPS. Employment reacts to motherhood more or less like a step-function, with little evidence of pre-trends and generally no sign of recovery in the ten-year post-period we observe.[11] Note that given the longer sample periods covered in our U.S. data, we can extend the post-period an additional five years.

[11] With our data we cannot observe persistence in the *very* long run, and recent evidence suggests that mothers may make up some of their lost prime-age employment by delaying retirement (see, e.g., Goldin and Katz, 2016).

It is worth emphasizing that Figure 3.3 and our other event-study figures plot event-study coefficients (how employment changes relative to $\tau = -1$, conditional on all the controls in equation 3.1) and thus except for $\tau = -1$ (which is, by construction, normalized to the raw mean in that year) raw means cannot be read off the figure. For example, for each of our four panel datasets, female employment is on a positive, secular trend, which we capture with the year fixed effects. So, in raw levels, women *do* increasingly return to work as time passes after the birth of their first child, but not any faster than the year fixed effects and other controls would predict. Appendix Figure A.2 demonstrates the same point in a slightly different manner, using childless women as the counterfactual instead of modeling the counterfactual with year and age fixed effects as we do in Figures 3.3 and 3.4. We use the NLSY and plot employment for childless women, generating a "fake event" of childbirth (see the figure notes for more detail). We then compare this series to employment outcomes for women who become mothers (where both series are simply raw means in event-time). For the most part, the childless-women series is increasing slightly each year, as is the series for mothers, except for the large decrease the year of childbirth.[12] In the paper, for the sake of exposition we will often refer to an outcome as rising or declining upon motherhood. However, such a statement rarely refers to raw levels, but instead to changes conditional on the controls in the event-study specification.

Returning to Figures 3.3 and 3.4, we see some differences in levels and magnitudes, though given that the exact definition of employment varies slightly among our four datasets, comparisons should be made cautiously. U.S. women are starting, in the year before parenthood, from a somewhat lower employment baseline than women in the BHPS. This difference seems mainly due to more U.S. women being students in the years right before motherhood, consistent with their being younger at the time of first birth (and thus closer to their high school or early post-secondary years)

[12]The same exercise for the other three panel datasets is available upon request.

and attending school longer than their British counterparts. Among the U.S. data sources, ordering the results from the oldest to the most recent birth cohorts, we see the "mommy effect" decline somewhat over time: it is -0.42 in the NLSW68 (average birth cohort of 1951), -0.30 in the NLSY79 (average birth cohort of 1962) and -0.27 in the PSID (average birth cohort of 1967). While the decline between the NLSW68 and NLSY79 is large enough so that the difference is unlikely a mere artifact of employment definitions, we find it harder to interpret the smaller difference between the NLSY79 and the PSID. We take away that even for our most modern cohorts, the effect size remains large.

How do our results compare to existing work? Figures 3.3 and 3.4 very closely parallel the results in Kleven *et al.* (2018), which uses Danish register data drawing from the 1955-1973 birth cohorts. Like us, they find little evidence of pre-birth pre-trends, a large effect the year of the birth, and little if any long-run recovery. Our effect sizes, however, are much larger—Danish mothers are roughly twelve percentage points less likely to work than they were before motherhood, whereas the decline for British women is roughly forty percentage points and for American women twenty-five to forty percentage points, depending on the cohort. In general, Denmark has a much higher female labor force participation rate than the UK or US (and in fact, higher than essentially any other country), and comparing our results to those using Danish data suggests that a large part of the difference likely comes from women's different responses to motherhood. Smaller effect sizes in Denmark may be explained by the country's more generous parental-leave and child-care policy. In addition to covering 52 weeks of parental leave, the Danish government guarantees all children between the ages of six months and five years a spot in heavily subsidized child-care facilities. By contrast, U.K. and especially U.S. parental leave is much less generous, a point to which we briefly return in Section 7.

3.4.2 Robustness and related results

As discussed earlier, we present a standard set of robustness tests across our various datasets. These checks include adding, as controls, a person-specific linear time trend, dropping as controls the individual fixed effects, adding in a control group of childless adults, and including or excluding "agers-in," children of original sample households who become eligible for interviews as they come of age.[13] In no case do these changes have an appreciable effect on the estimated difference-in-difference "mommy effect" for employment, as seen in Appendix Tables A.2 and A.3. Appendix Table A.4 also shows, as we would expect, that employment effects for men in the UK and US are very small (and of mixed signs).

Appendix Table A.5 shows that, in all four datasets, the effect of a second child of maternal employment is statistically and economically significant, though also smaller in magnitude than the first (about one-fourth the size of the initial "mommy effect"). The table also shows that both effects can be separately distinguished from each other even after including a person-specific linear time trend.

Finally, one concern with the event-study graphs is that, as we do not impose that every observation helps identify every event-study coefficient, respondents who mostly help identify the pre-period coefficients (e.g., those who have children at older ages) may differ in their reactions to motherhood from those who mostly help identify post-period coefficients. As a check on this concern, for each dataset, we split respondents based on the share of their observations that come from the pre-period and run separate event studies for the two groups (recall that, by construction, all observations in our preferred sample have at least one observation before and after the event). We show results for the BHPS (which is our most imbalanced event study) in Appendix Figure A.3 and the shape of the two event studies are nearly identical. This same

[13] While the control group of childless adults exhibit no variation in the *Post-baby* variable, they help identify the year and age fixed effects.

exercise for our other datasets and outcomes also shows little difference. Results available upon request.

3.4.3 Heterogeneous employment responses to motherhood

So far, we have shown that women in the US and UK exhibit a sudden, large, persistent and robust drop in employment, coinciding with the birth of their first child. In this subsection, we examine whether certain person-level characteristics predict the magnitude of this response. We are especially interested in heterogeneity by characteristics that are often assumed to increase women's attachment to the labor force (education, delaying motherhood, and having had a mother who herself worked) or relate to her expectations about the employment costs of motherhood (her pre-baby attitudes toward working mothers or her teenage expectations of whether she would work in her thirties).[14]

In the interest of space, we will generally examine heterogeneous responses to motherhood in a difference-in-differences as opposed to an event-study framework. That is, we take our difference-in-difference equation (based, again, on equation 3.1) and add an interaction between $\textit{Post-baby}_{it}$ and the binary, person-level variable of interest X_i (e.g., having had a working mother yourself). For each of our X variables of interest, the panels of Figure 3.5 show the point-estimate and confidence interval for the "mommy" effect for women with $X_i = 0$ (labeled "No" in the figure) versus $X_i = 1$ (labeled "Yes") and report as well the difference and its level of significance.

[14]Examples of recent studies that examine the impact of these characteristics on women's employment and wages include: Fitzenberger *et al.* (2013) and Wilde *et al.* (2010) (for education); Miller (2011), Herr (2015), and Herr (forthcoming) (for delaying motherhood); Olivetti *et al.* (2016) and McGinn *et al.* (2015) (for own-mother's employment). While we do not focus on them as they do not directly relate to predicted labor-force attachment or expectations, we also examine differences by race (in the US only, as the share of non-whites is too small in the BHPS), marital status, and gender of the first-born child. Blacks have smaller mommy effects (mostly driven by lower pre-baby employment levels) than other respondents. The mommy effect does not differ by marital status or gender of the first child for any of our panel datasets. Given that roughly forty percent of our respondents are unmarried at the time of first birth, we do not focus on heterogeneity by characteristics of the partner.

The heterogeneous employment responses to motherhood along the dimensions we study are remarkably similar across datasets. In panel (a) of Figure 3.5, we allow the mommy effect to vary by education. In the BHPS, women with a college degree have a statistically significant but modestly smaller (roughly eight percentage points) employment decline than their less educated counterparts. In the NLSW68, the "mommy effect" is roughly six percentage points smaller for women with a college degree relative to their less educated peers; in the NLSY79 it is ten points smaller; and in the PSID thirteen points smaller. All of the differences in the U.S. data are statistically significant. However, it is worth emphasizing that in all four datasets, even more educated women in the US and the UK exhibit very economically and statistically significant declines in employment upon motherhood.

In panel (b), we see that delaying childbirth (which we define in each dataset as having a first child at an age above the female median) offers no such protection, and in fact even the sign of the effect varies across data sources (though magnitudes are uniformly small and insignificant). In panel (c), we find similar "non-results" for having yourself had a working mother. This result contrasts with Kleven *et al.* (2018), who find an intergenerational effect of working motherhood.[15]

While we can look at these three attributes in all four of our datasets, the remaining panels of Figure 3.5 do not include all four datasets due to differences in variable availability. For the BHPS, NLSW68 and NLSY79, we can stratify by pre-baby views about how women should divide their time between work and family, and for lack of better terminology we call those women who say women (and mothers) should work as having "pro-work" views. (Note that we use this gender-role variable as an outcome in the next section and describe it in more detail then.) To generate two groups, we

[15]We speculate that the key difference is that Kleven *et al.* (2018) are able to better control for family wealth during childhood. For example, in our data, having had a working mother could correlate with having grown up in a poorer family where the mother had to work. If poorer families instill more conservative gender-role attitudes in their children, then this omitted variable would dampen the observed intergenerational correlation of labor supply between mothers and daughters.

split the sample based on the pre-baby median of this variable. Perhaps not surprising, women who have (pre-motherhood) stronger beliefs that women should work also exhibit smaller "mommy effects," though this result is not significant at conventional levels in the NLSW68. In the NLSW68 and the NLSY79 we can stratify observations by whether, in their teen years, women predicted they would be in the labor force at age 35. These women tend to have significantly smaller "mommy effects," but even women who predicted they would work in their thirties are experiencing large declines in employment upon motherhood.[16]

3.5 Do women anticipate the "mommy effect" in the short-run?

Section 3.4 provided a variety of evidence showing the large, negative employment effects of motherhood. In the UK (US), women's employment rates decline 38 (28–40) percentage points upon motherhood, with no evidence of recovery. To put this decline in perspective, in the most recent CPS, employment rates are 23 percentage points lower for 65-69 year-olds than for 60-64 year-olds.[17] Thus, our estimated "mommy effect" is comparable to a retirement effect.

While a large literature studies whether households fully anticipate and prepare for the effects of retirement, the question of whether women anticipate the employment effects of motherhood has received relatively less attention. As we will discuss in more detail in Section 3.7, dynamic labor supply models simply assume that women accurately anticipate these effects and rationally incorporate them into human capital and fertility decisions.[18] This section provides several pieces of evidence suggesting

[16] We defer discussion of the "family-friendly" occupation results in panel (f) to Section 3.7.

[17] See https://www.bls.gov/cps/cpsaat03.pdf.

[18] An important exception is Fernández (2013), who explicitly models how women form beliefs over the effect of working on family life, using signals they inherit from their own mothers as well as common signals embodied in the current female labor force participation rate. While we adopt a

that women, in the short- to medium-run, are systematically underestimating the effect of motherhood on employment. Section 3.6 takes up the question of the accuracy of their expectations in the longer run.

3.5.1 Effect of parenthood on views toward gender roles and female employment

We begin by exploring whether women's beliefs about the proper balance between family responsibilities versus market work change upon having a child. If, as we argue, women underestimate the costs of motherhood, then the event of the first birth serves as an information shock and thus beliefs should be updated.

To measure how beliefs change, we make use of six questions in the BHPS. These questions are in the form of statements about the proper role of women in the workplace versus the home, and respondents are asked to state their agreement with each statement on a 1–5 scale (e.g., "All in all, family life suffers when the woman has a full time job," "Both the husband and wife should contribute to the household income"). If needed, we reverse the order of the responses so that each question is increasing in the pro-female-employment direction. We then standardize each of these variables and sum them. Appendix Table A.6 provides summary statistics for this "omnibus" measure of summed standardized variables as well as each component (and the wording of each component statement). Note that, in contrast to the employment question, which is asked every year, the BHPS asks these six gender-role questions every other year, so we expect point-estimates to be noisier given the smaller sample.

We refer to this omnibus variable as our *index of gender-roles regarding work and family*, or simply *gender-roles index*, for expositional ease. To get a sense of

much more empirical approach, we share her view that it is legitimately hard for women to predict the costs of working post-motherhood. As she writes: "It is not an exaggeration to state that, throughout the last century, the consequences of women's (market) work have been a subject of great contention and uncertainty."

magnitudes, note that women are generally 0.77 units more in the pro-work (that is, agreeing that it is good for women to work) direction than are men.

This index has strengths and weaknesses. A strength is that all the questions refer to women's roles as workers versus mothers and wives. One could imagine instead questions about gender stereotypes, but not about employment (e.g., "On average, women and men have equal innate intelligence"), which would be less germane to our hypothesis. Another strength is that the question is abstract and does not specifically ask about the respondent's *own* household, as such a question would not lend itself to event-study analysis. The BHPS statements are about households *in general* and thus respondents without children can still answer these questions.

A weakness is one that arises for any type of subjective response: it is impossible to know exactly what the respondent means. But because statements in the BHPS are not about one's own situation but about the appropriate responsibilities for men and women generally, in principle a woman who perfectly anticipates the costs of motherhood should give the same response before and after becoming a parent. By contrast, a woman who overestimated (underestimated) the employment costs of motherhood should change her answer in a pro- (anti-) work direction once she learns how easy (hard) motherhood actually is and thus how realistic or desirable it is for mothers to work while raising children. Our implicit assumption is that respondents are answering this question about households in general, but are using their own experiences to inform these answers.

Main results

Figure 3.6 shows the results of our event-study analysis of gender-role attitudes. The birth of the first child is associated with women moving 0.92 units in the *anti-work* direction on our omnibus gender-roles index. The effect is not only statistically but economically significant—in fact, it is slightly larger than the male-female difference.

Like the employment effects, there is little if any pre-trend, though as expected the results are noisier, with larger standard errors. The effects of childbirth on our gender-roles index take about two years to be fully realized: attitudes become somewhat more anti-work the year of birth and even more so the year after. Beyond that point, they stabilize and show no evidence of recovery. Appendix Figure A.5 shows no statistically significant effect of fatherhood, though, if anything, men also also adopt more anti-female-employment views upon parenthood as measured by our index.

As we did with the employment event-study results, we can use childless women to generate a counterfactual instead of modeling it with year and age fixed effects. Appendix Figure A.4 performs this exercise. Outside of perhaps a slight, smooth movement in the anti-work direction, childless women for the most part do not change their gender-role norms over time, whereas the change around the time of childbirth is evident for mothers.

As noted, this omnibus index is composed of six underlying questions. In Appendix Figure A.6 we show how women's views evolve for each component of our omnibus measure, returning to our standard specification. Of the six components, five show statistically significant movement in the anti-work direction and one ("pre-school child suffers if mother works") shows an insignificant movement in the pro-work direction.[19] Some of the individual components of the gender-roles index have a modest pre-trend, but we show in Appendix Table A.7 that all five of the outcomes with a significant "mommy" effect retain their sign and significance even after adding a person-specific linear trend.

We can in principle replicate this analysis in the NLSY79, though we emphasize an important caveat. The battery of gender-role questions are asked only four times: in 1979, 1982, 1987, and 2004. The fact that the questions span a long period means

[19] We have no intuition for why this question moves in the opposite direction of the others, but do note it is the only question that refers to *pre-school* children, so is somewhat misaligned in terms of our event-study and difference-and-difference assumption that the event is the birth itself.

we can observe essentially the entire sample over their child-bearing years, as almost everyone in our NLSY79 sample who will have a child does so between 1979 and 2004. However, the infrequency of the questions means that the event study is highly imbalanced and there are relatively few (100-300) non-missing observations underlying each coefficient that we plot.[20] The wording of and summary stats for the NLSY79 questions can be found in Appendix Table A.8. In general, the questions in the BHPS and NLSY79 are quite similar, and ask agreement with statements relating to two broad themes: (1) the roles a husband and a wife should adopt with regard to splitting paid work vs. housework and (2) the impact on children (and family) if a woman works. Fortunately for us, both sets of questions focus on *employment* in particular, rather than gender more broadly.

Figure 3.7 displays event-time coefficients that roughly support the results from the BHPS: that motherhood is associated with a decline in support for the idea that female employment does not detract from family life. It is important to note that, as expected, standard errors are substantially larger and that there is a non-trivial, negative pre-trend in the years before motherhood.

As with the employment results, we subject the gender-role results to our usual battery of robustness checks (adding a linear time trend, using additional controls group of non-parents, etc.). For each robustness check and for both datasets, the effect of the first child's birth on moving the mother in the anti-work direction remains statistically significant with a very similar point-estimate to the baseline result, as seen in Appendix Tables A.9 and A.10. The effect of a second child is an additional move in the same direction, large and highly signifiant in the BHPS. This effect is negative but insignificant in the NLSY79, though given the gender-role questions are only asked four times in the NLSY79, separately identifying the effect of both a first

[20] Although the NLSW68 also contains similar questions, which we exploit in panel (f) of Figure Figure 3.5, they are not asked in a consistent manner over time so we cannot replicate the event-study analysis in the NLSY79.

and second birth may be asking a lot of the data. The results for men are similar in that they move in the same direction as women do on the gender-role index after parenthood, but effects are smaller and not always statistically significant.

Heterogeneous effects

We examine heterogeneity in how attitudes change in response to motherhood along the same dimensions we earlier examined for employment. Panels (a) through (c) of Figure 3.8 show that, for both the BHPS and NLSY79, all of the factors that we hypothesized would make mothers remain attached to the labor force (advanced education, delaying childbirth, and having had a working mother), produce *larger* movements in the anti-work direction upon motherhood, though two of the six estimated differences are not significant at conventional levels. Similarly, panel (d) compares women with relatively more pro-work gender-role attitudes pre-motherhood to those with more anti-work views. Obviously, this regression is complicated to interpret because we are stratifying on the dependent variable, but we include it nonetheless to remain parallel to Figure 3.5. In fact, the entire anti-work shift in attitudes is driven by women who were more pro-work pre-motherhood. In the NLSY79, we can also stratify the sample on whether they predicted at age 18 that they would be working at age 35, and, again we see larger shifts for those who assumed they would be working.

Note that, for both datasets, heterogeneity in the response of gender-role norms to motherhood generally takes the *opposite* sign of the heterogeneity in the response of employment. For the employment outcome, we found that education had the effect of blunting the mommy effect. By contrast, educated women appear the most affected by motherhood in terms of their answers to the gender-role index questions. The same pattern holds (i.e., smaller employment effects but in fact larger attitudinal shifts) for having less traditional views on the gender-role index pre-baby and (in the NLSY79) having predicted working at age 35. These conflicting patterns suggest that,

importantly, the gender-role results *cannot* be explained by simple models of cognitive dissonance, whereby individuals change their beliefs about the proper balance between family and work to match their own changing patterns of behavior. We return to these opposing patterns in Section 3.7.1.

3.5.2 Evidence from retrospective questions

We have so far shown robust evidence that in relatively recent cohorts in both the UK and the US (recall, average birth cohorts of 1968 and 1961, respectively), women adopt more traditional views about the appropriate balance between work and family responsibilities upon becoming mothers. We interpret these results as demonstrating that, pre-motherhood, women on average underestimate the employment costs of motherhood. This lack of anticipation appears concentrated among more educated women.

We can, in fact, look more directly at whether parents are underestimating the costs of child-rearing. In 1997, 2001, 2007 and 2013 the PSID asks subjects with children if parenthood is harder than they thought it would be (the average birth cohort among respondents to this question is 1968, so similar to those we use to estimate the gender-norms event study). Obviously we cannot investigate how this question evolves in event time, as it is not asked of individuals *until* they become parents. We can, however, determine whether educated women are more likely to agree with this statement than their less educated counterparts, which would be consistent with our reading that their larger change in responses to the gender-role questions in the previous subsection represents a greater underestimate of the costs of motherhood.

We examine how both mothers and fathers answer this question, stratifying by education. Table 3.2 shows results from regressing, separately by gender, an indicator variable for parenting being "harder than I thought" on a college-degree indicator and

a vector of standard covariates (which we vary to probe robustness). Before discussing the regression coefficients, it is interesting to note the large differences in the mean of the dependent variable by gender: whereas 52 percent of mothers with children age six and under report having underestimated how hard parenting would be, only 37 percent of fathers with similarly aged children respond the same way.[21] Most striking, in col. (1) we see that those mothers with a college degree are 12 percentage points more likely to agree than their less educated counterparts, an effect robust to adding controls (col. 2). It remains significant (though smaller) when we include mothers of older children (col. 3). For men (cols. 4–6), education is not a significant predictor of agreement with this sentiment.

3.5.3 Do women correctly predict the effect of motherhood on labor force participation in the short run?

In the previous subsection, we argued that, if women do not fully anticipate the costs of motherhood, then the birth of a child is an important informational shock, and we should see the updating of beliefs. And, indeed, we find evidence supporting this claim in both U.K. and U.S. data.

In this subsection, we more directly examine anticipation of employment effects, returning to our usual event-study framework and using as an outcome variable a BHPS question asking respondents whether they foresee giving up paid work in the next twelve months. *Unfortunately, the question is only asked of individuals who currently have a paid job.* Given the evidence in Figure 3.3 that women in the BHPS substantially reduce their labor supply upon motherhood and the evidence in Fig-

[21]Respondents were asked to rank the truth of the statement "Being a parent is harder than I thought it would be" on a scale of (1) not at all true to (5) completely true. Responses for our regression sample are as follows: 16 percent of women and 22 percent of men select (1) "not at all true;" 11 percent of women and 21 percent of men select (2); 22 percent of women and 21 percent of men select (3), 18 percent of women and 20 percent of men select (4), and 33 percent of women and 17 percent of men select (5), "completely true."

ure 3.5 that this reduction is mediated by covariates such as education, the sample who remains post-motherhood to answer this question is highly selected. Ideally, a question such as "Do you expect to be in paid work twelve months from now" would have been posed to all respondents (so, for those without employment, would refer to finding or returning to a job). As such, we view these results as more suggestive and do not make them the core of our argument.

Sample-selection concerns notwithstanding, we conduct an event-study analysis using two related outcomes: the share of women who, in year t, will leave the labor market in year $t+1$; and the share of women who, in year t, predicted they would leave the labor market in year $t+1$.[22] We include only women who work in year t, so both outcomes are affected by the same selection biases. Figure 3.9 shows that, pre-motherhood, these two series closely track each other and are close to zero: almost no working women predict they will drop out of the labor force the following year, and essentially none do. However, the two series start to diverge post-childbirth, with actual rates of giving up paid work increasing roughly 17 percentage points, while the share of women who *predict* they will leave increases by less than half this amount. Interestingly, mothers continue to overestimate the probability they continue working even several years after the birth of the first child, perhaps due to underestimating the effects of future children (recall that we find a second-child effect for both employment and our gender-role index outcomes, Section 3.4.2 and Section 3.5.1, respectively) or underestimating the demands of the first child at older ages. Appendix Table A.11 reports the regression analogue of this figure, using a dependent variable that is equal to one if the respondent expected to work this year but was wrong, and subjects the result to our usual battery of robustness checks. For men (columns (7) and (8)), we find no statistically significant change post-baby.

[22]Note that for these purposes we define the *event* of child-birth as corresponding to the prediction in year $t-1$, so the $t=0$ observation corresponds to the prediction made in the year before childbirth ($t=-1$), realized in the year of childbirth.

In summary, post-motherhood, there is substantial increase in the likelihood of a working mother in year t dropping out of the labor force in year $t+1$. However, less than half of these departures are predicted by the women even one year earlier. We emphasize that sample selection is an issue, but one point in defense of this exercise is that, given that a large share of women drop out of the labor force (and thus the universe for this question), we expect that we differentially observe in the post-period those women most attached to the labor force and thus who would be the most likely to remain in any given year post-motherhood. As such, we feel it is likely that sample selection leads us to understate the tendency of mothers who report that they will remain in the labor force when in fact they do not.

3.6 Do young women accurately forecast their future labor supply in the longer run?

In the previous section, we presented evidence that women underestimate the costs of motherhood in the relatively short-run. In this section we take a longer-run view. Specifically, we ask whether, at the time they are making key human-capital and early career decisions (i.e., their late high school and college years), young women are able to forecast their prime-age employment behavior.

3.6.1 Data and variable definitions

Some of the analysis in this section makes use of the NLSW68 and the NLSY79, and some relies on data we have yet to introduce. Summary statistics for these supplementary data sources can be found in Appendix Table A.12 and more details on the data can be found in Appendix A.2.5. We provide a brief overview below.

Monitoring the Future (MTF) is an ongoing study of the behaviors, attitudes, and values of a nationally representative sample of U.S. students. We examine twelfth-

graders' responses to a question on the type of work they expect they will be doing at age thirty. Specifically, the question asks, *"What kind of work do you think you will be doing when you are 30 years old? Mark the one that comes closest to what you expect to be doing."* There are fifteen occupational categories from which students can choose, with "home-maker" as one option. See Appendix A.2.5 for a list of the occupations provided by the MTF for this question (as well as the list of occupations offered to respondents in the other data sources we describe in this subsection).

The NLSY79 asks respondents what they plan to be doing at age 35. Unlike in the MTF, respondents can only choose among four options: "working at current job," "working at a different job," "married, raising a family," and "other." Respondents who indicate that they plan to be "married, raising a family" are *then asked a follow-up question* about whether they also plan to be working outside the home at age 35. We define expected home-maker in the NLSY79 to be those who indicate planning to be "married and/or raising a family" *and* not working outside the home at age 35.

For years prior to 1977, we can in principle rely on the NLSW68. However, we are concerned that the NLSW68 may *overstate* the share being categorized as "home-maker," since respondents are forced to choose between motherhood and work. Specifically, respondents are asked the *first* question from the NLSY79 but then those who choose "married, raising a family" are *not* asked the follow-up question on whether they also plan to work outside of the home.

We thus supplement the pre-1977 period with two additional, cross-sectional surveys: a 1961 Purdue Opinion Panel of high school students grades 9–12 and a 1961 Gallup opinion poll of young high school, college, and non students. The Purdue Opinion Panel asks "What kind of job do you expect to have 20 years from now?" and respondents can choose between twelve possible occupations, including "housewife." Respondents in the Gallup poll were instead asked the open-ended question "What do you expect to be doing when you are 40 years old?" Responses were or-

ganized into 18 categories. We use the "house wife," "home maker," "house work," and "raising children" category as our definition of expected homemaker. In both surveys, we restrict our data to female high school seniors, to match the MTF.[23]

How do we compare the expectations of teenagers to their actual behavior once they are in their thirties? For the longitudinal surveys, we simply compare the expectations and realizations of the same group of individuals (so, we look only at those individuals who *both* answer the expectations question in their teens *and* report their realized employment status in their mid-thirties).[24] For the cross-sectional surveys (the MTF, and the Purdue and Gallup surveys), we compare respondents to their "synthetic selves" in the CPS 12-20 years later (this twelve-to-twenty-year range accounts for the fact that the MTF asks about expectations at age 30, twelve years after age 18; the NLSY79 at age 35, seventeen years after; the Purdue "twenty years later;" and the Gallup at age forty, twenty-two years later). That is, we compare answers regarding expectations at age 18 to realizations averaged over the ages 30 to 40.[25]

3.6.2 Results

Figure 3.10 depicts several series of expectations and realizations. We describe these results chronologically. The earliest expectation results we have are from the Purdue and Gallup surveys, both from 1961. Reassuringly, results from these surveys agree with each other: fifty-five to sixty percent of women in both surveys predict that

[23]A potential concern is that the term "housewife" or "homemaker" is dated and has been replaced by the more modern "stay-at-home mom," and that this antiquated language potentially lowers the share of women who would see themselves in this role. Using Google Ngrams, we find that both "housewife" and "homemaker" are still used far more often than "stay-at-home mom" through 2008 (the last year NGrams provides data). Moreover, the term "stay-at-home mom" is almost never used until the late 1990s.

[24]In fact, the MTF does appear to include a longitudinal component, but the data are not shared and in fact we can find few citations that make use of these data. Not even codebooks appear to be published.

[25]In previous versions, we matched each expectation to the exact future age to which it referred. Results are nearly identical, though far more cluttered visually.

in their late thirties they will be home-makers. In general, these shares are quite similar to those based on eighteen-year-olds in the NLSW68 from later in the decade. Consistent with Goldin (1990) and Goldin (2006b), we find that, in the NLSW68, roughly two-thirds of young women in 1968 expect to be home-makers at age 35, with this share falling over the following two years.[26]

Comparing these expectations to the realized labor supply of women twelve to twenty years later, we see that eighteen-year-old women in the 1960s vastly underestimated their future labor supply (again consistent with Goldin's earlier work). Comparing their expectations to the behavior of their future, synthetic selves in the CPS, women in the Purdue and Gallup surveys overestimate the probability they would be housewives by about fifty percent. As the actual housewife share declines rapidly over the next ten years, the women in the NLSW68 in fact make more severe (by a factor of two) overestimates of their future housewife status. (This result holds whether we compare the NLSW68 women to their future selves in later waves of the study or to their synthetic selves in the CPS.)

Sometime between the high school class of 1968 and 1978, young women's tendency to underestimate their future labor supply disappears and in fact reverses. The share of the MTF class of 1978 who predicts being a housewife at age thirty is roughly ten percent, a dramatic decline from their counterparts in the 1960s and early 1970s. Moreover, while the 1960s high school classes were *overestimating* the probability of being a housewife, from 1978 onward, every high school class in our data underestimates it. For the late 1970s high school classes (when we can compare the MTF and the NLSY79, as their samples overlap in terms of birth cohorts during this period) the magnitude of the underestimate is very similar whether we compare

[26]Our panel data sources, the NLSW68 and the NLSY79, ask its expectations questions several times. The points plotted in Figure 3.10, however, use only the data from the year a respondent turns 18, to match the MTF. A small concern is that some of these 18 year-olds will have seen the question in previous waves of the survey. Readers concerned with this potential anchoring bias can simply ignore the points after 1968 and after 1979 for the NLSW68 and the NLSY79, respectively.

MTF respondents to their synthetic selves in the CPS or the NLSY79 respondents to their own future selves: around ten percent of respondents predict they will be housewives, but in reality just under twenty percent will be. Since about 1990, less than two percent of our MTF sample predicts they will be housewives at age thirty, whereas the actual housewife share in the CPS has not declined appreciably since 1980 (it has bounced between 15 and 18 percent).[27]

One natural explanation is that more recent cohorts of young women are underestimating their *fertility*—perhaps they do not expect to have children (or expect to have fewer than they actually will). That is, instead of underestimating the employment costs of motherhood, they simply underestimate the probability of motherhood itself. In fact, Appendix Figure A.7 shows that, if anything, women in the MTF overestimate their future fertility. As such, they plan on being mothers, but, at the same time, do not connect fertility with home-making.

Taken together, these results indicate that there has been a sharp reversal in the accuracy of young women's predictions of their future labor supply. Throughout the 1960s and early 1970s, we find that young women are largely *overestimating* the probability that they will be home-makers. From 1978 through the early 21st century, we instead find that young women are *increasingly underestimating* this probability. As such, since the late 1970s, women who are on the cusp of making key human capital investment decisions appear to substantially overestimate their future attachment to the labor market.

It would be interesting to examine whether women's expectations become more accurate in their twenties, when many of them will be in college, post-baccalaureate

[27]This result that young women seem unconcerned or unaware about the future challenges of child-reading echos recent work by Lordan and Pischke (2016). In a survey experiment, they ask 18-year-olds to make pairwise choices between possible careers, and later to explain in an open-ended format why they made those choices. No students, male or female, mention the need for flexibility or childcare, but focus instead on the attributes of the job. The disregard for the future challenges of childcare suggested by their and our results is consistent with speculation in Bertrand (2018): "It is also possible that women today more than in the past believe (rationally or not) that they can 'have it all.' "

studies, or making early career decisions. While we know of no dataset that would allow quantitative analysis of this question, recent ethnographic evidence suggests that, at least for high-achieving women in college, expectations remain quite poor. Schank and Wallace (2018) extensively interview 43 of their Northwestern sorority sisters who graduated, like they did, in the early 1990s. These women are surely not representative of all U.S. women, but they reflect high-achieving women who have made sizable investments in education.[28,29] Whereas all these women anticipated high-powered careers while in college and most left college for graduate degrees or jobs with long hours, only eight of the thirty-four who became mothers maintained full-time jobs. Fifteen moved to flexible or part-time work and eleven fully dropped out of the labor force.

Almost all the women who moved to flexible work or dropped out express surprise at the outcome. "I wasn't planning on staying home with my son after he was born," one mother said, "but once he was born, I realized something had to give and I needed to figure out what." Said another: "I never wanted to stay home with my kids, ever," but six weeks after the birth of her first child, the woman decided not to return to work. As the authors write, *only two* of the eleven mothers who dropped out of the labor force planned to do so before the actual birth.

In summary, this section has shown a variety of evidence consistent with the claim that, during the years they make their key human capital decisions, women do not fully anticipate the effects of motherhood on employment. Given how large the effect remains today and how common an occurrence motherhood is, our results so far raise

[28] As a reference, we compared the labor market outcomes of the Northwestern sorority sisters to women in our PSID sample, which, among our various U.S. datasets, most closely resembled the Northwestern graduates in age. Of the 34 sorority sisters who became mothers, 68% worked in some capacity post-baby. In our PSID sample, approximately 79% of comparable women (college graduate mothers in their early 40s) were working.

[29] As the authors point out, sororities at Northwestern are quite unlike typical sororities at other U.S. universities with "no hazing and a lot of studying." Women joined sororities at Northwestern to make friends and to gain a sense of belonging in a large school. As a result, the women in the sample were economically diverse and academically accomplished.

the question of *why* women would fail to anticipate these effects. We now turn to this question.

3.7 Why do women today underestimate the effects of motherhood?

Motherhood is associated with large and lasting declines in employment. What is perhaps most puzzling is that these large employment effects occur for cohorts of women who are far more prepared for the labor market than previous cohorts in terms of education levels (substantially higher than men's) and job experience (substantially higher than in previous cohorts, due to delayed childbirth). The large increase in investments are not surprising, as they are consistent with an enormous change in *expectations* about future labor supply, as we demonstrate in Section 3.6 (see also Goldin *et al.*, 2006b and Lee, 2018). And yet when motherhood arrives, these women exhibit substantial and lasting employment declines.

We hypothesize that the employment costs of motherhood are higher for current mothers than they would have predicted by projecting from past trends. Such an explanation has a number of attractive features. It explains why young women now (but not in earlier cohorts) overestimate the likelihood they will work at age thirty. It also explains why they have invested so much in education and have delayed childbirth despite, since roughly 1990, seeing no increase in labor force attachment (because they were making education decisions under the mistaken assumption of high, future labor force attachment, due to their underestimation of motherhood costs).

We begin this section with a simple model of women's educational investment decisions in the face of uncertain employment costs of motherhood. To bolster the credibility of this model, we first show that it yields, under quite general conditions, two of the more striking results from the paper: that, compared to their less ed-

ucated counterparts, college-educated women exhibit smaller *employment* effects of motherhood (recall Section 3.4.3, where we show that this result holds across all four of our datasets), but at the same time they are also more "surprised" by how hard motherhood actually is (recall Section 3.5.1, where we show their attitudes change more in the anti-work direction on work-family balance questions in both the BHPS and the NLSY79 and Section 3.5.2 where they are more likely to say in the PSID that parenthood is harder than they anticipated). *We show that this result holds whether or not the average cost of motherhood has increased or decreased relative to the previous generation.*

We then show that the model predicts some of our other key results—namely, that women *on average* update their attitudes in the anti-work direction upon motherhood and report that parenthood is harder than they expected —*only* when the costs of motherhood have increased on average across generations.

We complement this theory-based argument with a collage of empirical evidence that while the cost of motherhood fell during much of the twentieth century, by some important metrics, it has recently risen. Some of this evidence we take from existing literature and some, to the best of our knowledge, we provide here for the first time.

3.7.1 Modeling women's education and employment as a function of expected employment costs of motherhood

There have been several dynamic labor supply models from the macro literature that attempt to jointly explain women's fertility, human capital and employment decisions (or some subset of these decisions). Most of these models (e.g., Attanasio *et al.*, 2008; Blundell *et al.*, 2016; Adda *et al.*, 2017) assume that agents have perfect foresight of how children will affect labor supply. We present a much simpler, static model in which women predict the employment costs associated with their future children based on their observations of their own mother. In particular, when they are making

human capital decisions, they can over- or underestimate these future costs (which leads to under- and overestimates, respectively, of their future labor supply). In this sense, our model takes Fernández (2013) as inspiration, but imposes more functional form assumptions and is static (focusing on a single generation whose choices are impacted by the experiences of the previous generation).

Assumptions and set-up

Let utility $u(c, h)$ be a quasi-linear function of consumption c and labor h (for hours worked, say). Specifically, assume that

$$u(c, h) = c - \frac{h^{\gamma+1}}{\gamma + 1},$$

where $\gamma > 0$.

Women's consumption will be equal to market wages net of employment costs of motherhood (both per hour) times labor supply (in hours). Market wages are $\tilde{w} = w + \beta \cdot e$, where w is a base wage for those with higher education $e = 0$ and $\beta > 0$ is the hourly return to higher education $e = 1$. Gaining education $e = 1$ costs some "tuition" $\alpha > 0$, while $e = 0$ is free.

We view "employment costs of motherhood" very broadly, as any cost mothers endure during work. Recall, from the introduction, that these might include the per-hour cost of a nanny or day-care service, the emotional or psychic cost of being separated from the child while at work, the sense of guilt at underperforming as a mother or colleague, or simply lack of sleep due to working while also performing child-care activities. Actual employment costs for a given woman i are μ_i, but she predicts this cost with some error δ_i. In particular, she observes her mother's employment costs, $\mu_i + \delta_i$, but she only inherits the μ component. The δ component was her mother's "luck" (good or bad) and is *not* passed down to the daughter. This luck is the sum of a generational component and an idiosyncratic component. For example, her

own mother's idiosyncratic good luck might include having had a very understanding supervisor at work, or having parents or in-laws who live in close proximity and can provide free child care. Generational "good luck" could be, for example, giving birth in a moment with cheap, technological substitutes for child care. But the woman in our problem cannot distinguish between μ_i and δ_i and instead uses the sum as the best signal of her own, future employment costs of motherhood.[30]

Assume further that $\mu \sim U[0,1]$ and $\delta = \{\lambda - \epsilon, \lambda + \epsilon\}$ with equal probability, where λ and $\epsilon > 0$ are both constants. Note that λ can take negative values (for much of the analysis below, λ will drop out, but allowing $\mathbb{E}(\delta) \neq 0$ will be useful later). We assume μ and δ are independent. For simplicity, we further assume no taxes and that w is sufficiently large that $w - \mu - \delta$ is always greater than zero (so mothers' predicted effective wages are always positive).

We abstract away from many possible dimensions of heterogeneity. There is no variation in fertility in our model: women become mothers with probability one and we ignore issues related to delaying or timing childbirth. Unlike many models of human capital, we ignore any variation in "ability" and thus the effective return to and cost of education do not vary along this dimension.

The implicit timing of the model is as follows: A woman makes her education decisions assuming that her future employment costs of motherhood are $\mu + \delta$, but then will make her actual labor supply decisions after this uncertainty is resolved and her actual costs, μ, are revealed to her. While this timing is helpful to keep in mind, in fact the problem can be collapsed to a single decision: the woman's (only) problem is to optimally choose education $e \in \{0, 1\}$. Once she chooses e and once her true costs μ are revealed to her, her labor supply is given by a simple optimization problem and thus can be viewed as deterministic. Note that, as we are assuming

[30]Note also that women in our model are *naive* in the sense that they do not take into account that δ is drawn from a distribution and instead just assume employment costs will be $\mu + \delta$ with probability one.

away many other dimensions of heterogeneity, variation in optimal e comes entirely from variation in expected employment costs, $\mu_i + \delta_i$.

Working backward, the woman calculates her predicted utility for each $e \in \{0, 1\}$. This comparison requires calculating the optimal predicted labor supply $\hat{h}^*$, given her assumptions about employment costs, for each e and then determining which value for e yields higher utility. We describe key characteristics of both the predicted labor supply $\hat{h}^*$ and actual labor supply h^* in the Appendix.

3.7.2 Predictions yielded by the model

Claim 1. *Women who choose $e = 1$ will on average work more (post-baby).*

Proof. See Appendix. □

The intuition is relatively simple. Women who choose $e = 1$ have higher wages, and because the utility function is quasi-linear, there is no income effect of higher wages on labor supply. While some women with $e = 0$ will supply more labor than some women with $e = 1$ (because some women with $e = 1$ will have much higher employment costs than they had expected), on average those with $e = 1$ work more.

Note that to connect this result directly to the empirical work, we need to add slightly more detail to the timing of the model. We have effectively collapsed the model into a single period (women make their education decision based on expected costs of motherhood, and then instantaneously have a child and have their costs revealed to them). If we instead assume that women all have a (perhaps brief) pre-baby period where everyone works some max "full-time" hours (which is roughly what we see in our data), then the claim maps not only into educated women having higher post-baby labor supply, but also exhibiting smaller employment "mommy effects."

The next result states that, even though more educated women have smaller mommy effects with respect to employment, they are nonetheless the most "surprised" by the demands of motherhood.

Claim 2. *Women with* $e = 1$ *underestimate the costs of motherhood more than women with* $e = 0$*. That is,* $\mathbb{E}[\delta \mid e = 1] < \mathbb{E}[\delta \mid e = 0]$.

Proof. See Appendix. □

Recall that a "smaller" ("less positive" or "more negative") value for δ means that actual employment costs μ will be *greater* than expected costs $\mu + \delta$, thus leading to "surprise" at how hard child-rearing is. Intuitively, women who draw a low δ will, all else equal, be more likely to invest in $e = 1$, because they assume low employment costs of motherhood and thus high future labor supply.

As is made clear in the proofs, Claims 1 and 2 hold for any value of λ—that is, they hold if the previous generation had on average lower, higher or the same costs of motherhood relative to the current generation. The model thus yields our results on heterogeneity by education under fairly general assumptions, which we hope bolsters the credibility of the model for readers.

Now, we show that our *unconditional* results on belief-updating (i.e., that, *on average*, women's gender-role attitudes move in the anti-female-employment direction after baby and that they report parenthood being harder than they expected) hold *if and only if* the employment costs of motherhood have increased for the current generation relative to the previous.

Claim 3. $\mathbb{E}[\mu + \delta] < \mathbb{E}[\mu]$ *iff* $\lambda < 0$.

Proof. See appendix. □

In the next subsection, we discuss empirical evidence in support of our claim that maternal employment costs have on average increased. In fact, we describe

suggestive evidence for a stronger claim: that while they have decreased over much of the twentieth century, they have increased recently, which would likely have made it even harder for the current cohort of women to have predicted this evolution. Relative to our event-study evidence, the evidence in the subsection below is much less systematic and we view it more as a "collage of evidence," taken from a variety of sources. We see further work in this area to be an excellent path for future research.

3.7.3 Empirical evidence on the evolution of the costs of motherhood across cohorts

Several macroeconomists have modeled the *rise* of working mothers over much of the twentieth century as a function of *declining* costs of motherhood. Greenwood *et al.* (2005) argues that electrification liberated women from household chores, chronicling the dramatic decline in time cost associated with tasks such as laundry and cooking. Albanesi and Olivetti (2016) focus more on the biological aspects of motherhood. They show that the health costs of pregnancy and delivery have fallen dramatically since the 1920s, due to pharmaceutical innovations such as sulfa drugs and later antibiotics as well as the modernization of obstetric practices, all of which dramatically reduced mortality and morbidity associated with childbirth. Moreover, the quality of infant formula improved substantially over this time, minimizing the need for mothers to breastfeed.

Maternal employment remains a sensitive subject, suggesting that stigma costs play a role in mothers' decisions to work. Fernández *et al.* (2004) argue that World War II served as a negative shock to this stigma cost, as mothers were told it was their patriotic duty to work, and that this change in norms may have passed on to future generations.

Some recent work suggests that any decline in the cost of motherhood may have stalled or even reversed. Ramey and Ramey (2010) documents a dramatic increase

since the mid-1990s of the time-cost associated with motherhood in the US, especially for educated women (increases for fathers are similar proportionally, but from a substantially smaller base), trends that we reproduce in Figure 3.11. As late as 1994, mothers aged 25-34 spent roughly 11 hours a week on child-care activities, with little variation by mothers' education. By the early 2000s, college educated mothers spent 20-22 hours, while their less educated counterparts had increased their total to 16 hours.[31]

A major time cost associated with infants is breastfeeding. In a reversal of the trends highlighted by Albanesi and Olivetti (2016), mothers today are encouraged to breastfeed exclusively for at least six months and formula is deemed a distant, second-best option.[32] Indeed, mothers appear to follow this advice, at least in part. Figure 3.11 shows that breastfeeding rates decline among U.S. mothers in the 1950s and 1960s (we are unable to find U.K. data from this early period). By 1965, only twenty percent of U.S. mothers report ever having breastfed. But then in the mid-1970s, in both the US and the UK, the share of mothers nursing increases, before temporarily plateauing in the UK and modestly declining in the US during the 1980s. Since 1990, the share of mothers who report nursing their infants has exhibited a steady increase in both countries, especially in the US. Indeed, since 1990, this share has increased from approximately fifty to eighty percent. While we focus on "breastfeeding at all" because it is the measure we can best compare across time and countries, Appendix

[31]Borra and Sevilla (2015) document similar patterns in U.K. data.

[32]The U.K. National Health Services (NHS) recommend: "Exclusive breastfeeding (breast milk only) is recommended for around the first six months of your baby's life." Moreover, the NHS lists as a "myth" the claim that formula provides comparable benefits to breastmilk. The American Academy of Pediatricians official policy statement on breastfeeding reads: "The American Academy of Pediatrics reaffirms its recommendation of exclusive breastfeeding for about six months, followed by continued breastfeeding as complementary foods are introduced, with continuation of breastfeeding for 1 year or longer as mutually desired by mother and infant." See `https://www.nhs.uk/conditions/pregnancy-and-baby/your-breastfeeding-questions/` and `http://pediatrics.aappublications.org/content/129/3/e827.full#content-block`, respectively.

Figure A.8 shows that alternative measures follow the same U-shaped pattern over time.

Finally, we note that the cost of one of the most common substitutes for maternal inputs into child-rearing—professional child-care services—has grown steadily in real terms since the early 1980s. The final series in Figure 3.11 documents a 65 percent increase from the early 1980s to today.

Do women themselves perceive that the costs of motherhood have risen since they were themselves being parented? It is certainly possible that the increased costs we highlight above (time costs of child-rearing, prevalence of nursing, and monetary cost of child-care services) are offset in other ways. In fact, recent polls have asked women to make this intergenerational comparison.[33] Over 66 percent of current mothers say that being a mother today is harder than it was when they were children. Of course, mothers today are more likely to work than their own mothers were, and thus perhaps they feel they have it harder because they are being asked to both work and mother. But on the specific question of the demands of child-rearing, 56 percent of current mothers say they are more involved in their child's life than their own mothers were in theirs.

3.7.4 Is it harder to be a working mother today because workplaces have grown less family friendly?

We have so far provided evidence that the family component of work-family balance has become more difficult for mothers. But perhaps (instead or additionally) workplaces have become more hostile to mothers. While we have been implicitly assuming

[33]Specifically, we pool data on women aged 25-45 from following surveys: the ABC News Mother's Day Poll in 2006 (https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/4655) and the CBS Monthly National Poll from 2009 (https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/26947), 2010 (https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/31571), 2011 (https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/33962 and https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/33963), and 2013 (https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/34997).

that maternal employment costs are driven by the difficulty of finding substitutes for maternal child care, it could instead be the case that employment costs are driven by workplaces becoming hostile to women once they become mothers.

In many ways, workplaces are more family-friendly than in the past. In 1975, U.K. maternity leave policy entitled eligible women to twenty-nine weeks of leave, of which only six were paid (Gregg *et al.*, 2007). Since then, eligibility has expanded, as has the length of leave, currently at fifty-two weeks, of which twenty are paid (Glassdoor, 2016). In 1999, the government expanded subsidies and provision of daycare (Gregg *et al.*, 2007). By contrast, the US has much less generous support of working mothers, though it, too, has increased over time. At the national level, the Family and Medical Leave Act, passed in 1993, mandates that employers provide eligible workers with twelve weeks of job-protected *unpaid* leave (previously there had been no federal mandate for any amount of leave). Growth in paid parental leave has been driven primarily by private provision and state-level policies, with nine states currently having programs to fund paid leave (Gault *et al.*, 2014). Today, about fourteen percent of the civilian U.S. workforce has access to paid family leave (which includes maternity leave).[34]

Of course, there are important countervailing trends. Goldin (2014b) and Bertrand (2018) document the rise in the returns paid to long hours. Since women remain the chief providers of child care and tend to work fewer hours in the labor market, they are especially penalized in occupations that are characterized by these high and growing rewards to working long hours, many of which are highly remunerated.[35]

[34]See http://www.pewresearch.org/fact-tank/2017/03/23/access-to-paid-family-leave-varies-widely-across-employers-industries/.

[35]Cha and Weeden (2014) argue that the rising returns to long hours, coupled with the gender gap in the propensity to work overtime, plays a role in accounting for the slow convergence of the gender wage gap from 1979 to 2009. Indeed, there is abundant empirical evidence that women, particularly those with young children, place a higher value on flexibility at work and shorter work hours relative to men (Flabbi and Moro, 2012, Mas and Pallais, 2017 and Wiswall and Zafar, 2017). Not surprisingly, women respond to the challenges of combining career and family by disproportionately

While we cannot speak directly to whether workplaces on net have become more or less hostile to working mothers, we provide some suggestive evidence that the "work" component of work-family balance does not seem to be driving the employment "mommy effect."

The BHPS asks all current workers about their overall satisfaction with their job, as well as satisfaction on a number of specific aspects of their work (e.g., hours, prospects for promotion, pay). Of course, there is serious sample selection into who can answer this question because it is only asked of those who stay in the labor force (though our analysis includes person fixed effects, mitigating some concerns about composition bias), so we view this analysis as only suggestive. That (important) concern notwithstanding, we show in Figure 3.12 that women exhibit no decline in overall job satisfaction post-baby. Appendix Figure A.9 shows that, in fact, for two of the five components (total pay and hours worked), satisfaction increases significantly and for none does it decrease significantly.[36]

As a second piece of suggestive evidence, we return to the final panel of Figure 3.5. Here, we have divided women into whether their pre-baby occupations were "family-friendly," which we define as having a pre-baby occupation with an above the median share of women. In fact, we find no difference in the magnitude of the mommy effect on employment.[37] Again, while this evidence is indirect, it suggests that the "work" component of work-life balance is not the driving force behind the mommy effect.

sorting into family-friendly firms or public sector jobs after the birth of their first child (Hotz *et al.*, 2017 and Pertold-Gebicka *et al.*, 2016) and making occupational choices with these considerations in mind (Wasserman, 2018).

[36]While again only suggestive, ethnographic evidence supports our interpretation that it is not the work component of work-family balance that is causing motherhood to be "harder" than expected. After describing a long list of tasks associated with child care, one of the working mothers interviewed by Schank and Wallace (2018) says: "Then I go to work, where I can rest."

[37]We found this result surprising, and thus examined several different labor-supply outcomes (e.g., continuous hours instead of a binary employment indicator) and alternative definitions of "family-friendly" (e.g., percent of women in the occupation, percent of moms in the occupation, and percent of moms with children under age five in the occupation). In none of more than a dozen specifications did we find that having pre-parenthood a family-friendly occupation softens the employment effect of motherhood.

3.8 Conclusion

This paper contributes evidence to a growing literature showing that, despite huge increases in women's human capital investment, there remains a large and lasting negative effect of motherhood on labor-force attachment. We view our chief contribution, however, as demonstrating that women do not appear to anticipate this "mommy effect" on employment. They now substantially overestimate the likelihood they will be in the labor force during their chief child-rearing years, suggesting that they may be making education and other human capital investment decisions under mistaken assumptions.

We rationalize this behavior with a simple model. One implication of the model is that the costs of motherhood have risen in a manner today's cohort of mothers was likely unable to predict. In the final part of the paper, we add further evidence to recent work suggesting that such an increase, likely starting in the 1990s, indeed took place.

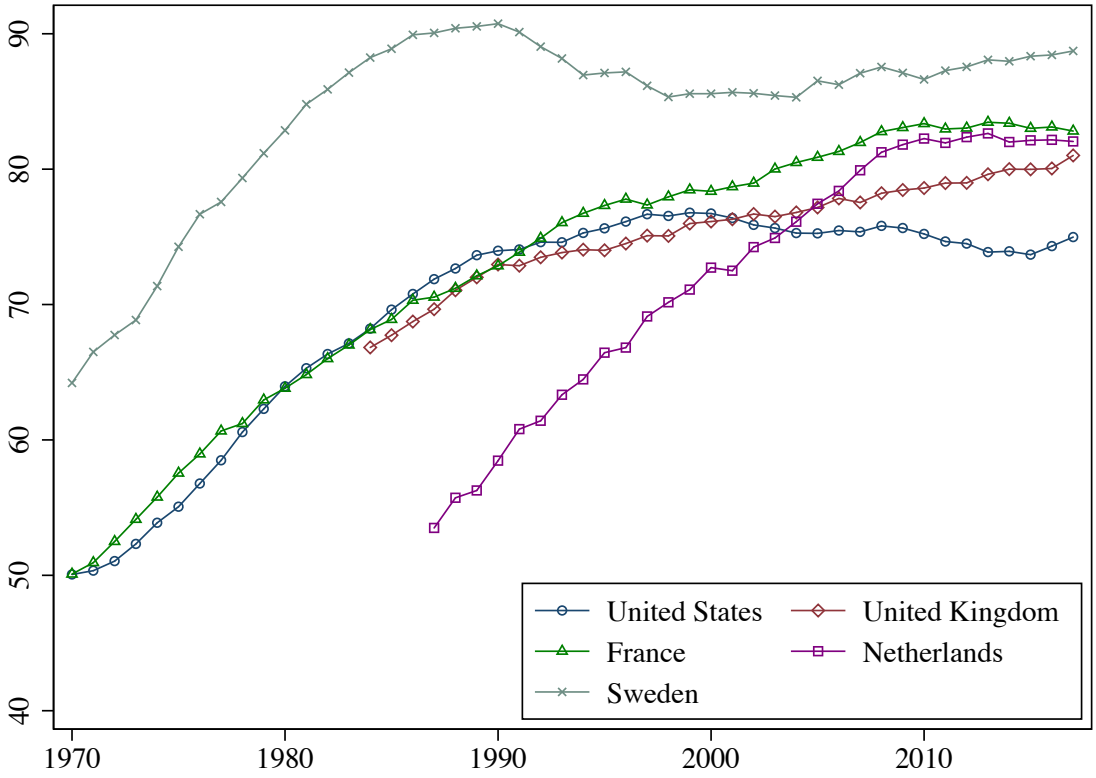
We offer two suggestions for future work. First, if current cohorts of women indeed overestimate their future labor supply (as we claim), are they in fact *over-investing* in education? Chen and Chevalier (2012) argue that female medical students systematically over-invest in education: given that female doctors work significantly fewer hours than male doctors, they argue that, from a purely pecuniary perspective, these students should have instead become physicians' assistants. Other papers have instead examined the *social* return of the rise of women's educational attainment. For example, Currie and Moretti (2003) show that exogenous increases in educational opportunities for adolescent women correlate with improved birth outcomes when these women become mothers. Yet other papers have emphasized the marriage market returns to women's education (Lefgren and McIntyre, 2006; Oreopoulos and Salvanes, 2011; Kaufmann *et al.*, 2015; Bruze, 2015), though it is important to note that these returns are private, not social, returns, as they are likely zero-sum. Related, women

may invest in education as insurance against the risk of divorce, which had been rising during the second half of the twentieth century (Bronson, 2014; DiPrete and Buchmann, 2006). Our work suggests that, unless the employment effects of motherhood decline in the future, non-labor market returns to female education are perhaps just as important to study as labor-market returns.

Second, throughout this paper we have taken as given that the employment costs of motherhood might change over time due to technological developments or changes in norms, but future work might focus on *why* they change.[38] Especially in the recent period, changes in maternal employment costs seem more driven by changing expectations and norms as opposed to changes in technology. On face, it is deeply puzzling that at a moment when women are more prepared than ever for long careers in the labor market, norms would change in a manner that encourages them to spend more time at home.

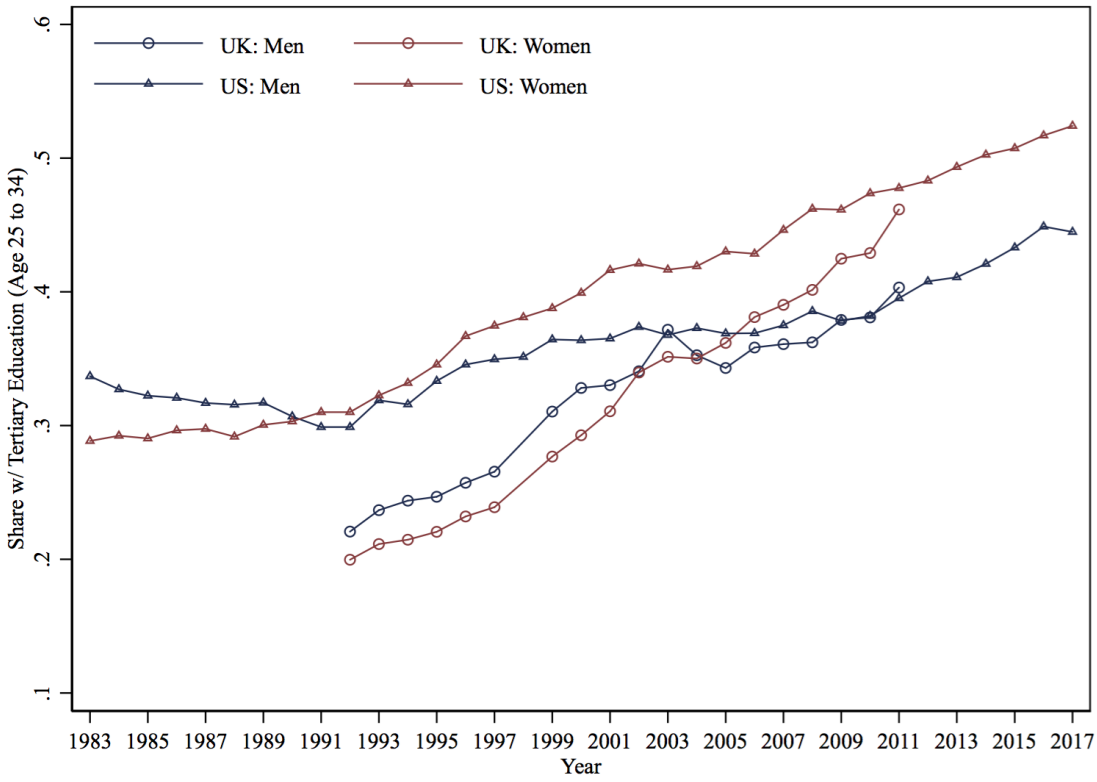
[38] A small number of recent papers suggest that the rise in involved parenting (and increase in time spent on children), particularly among educated parents, could be due to factors such as the increase in competition for slots in top colleges (Ramey and Ramey, 2010) and increases in inequality (Doepke and Zilibotti, 2017).

Figure 3.1: Female labor force participation rates in select countries (ages 25-54)



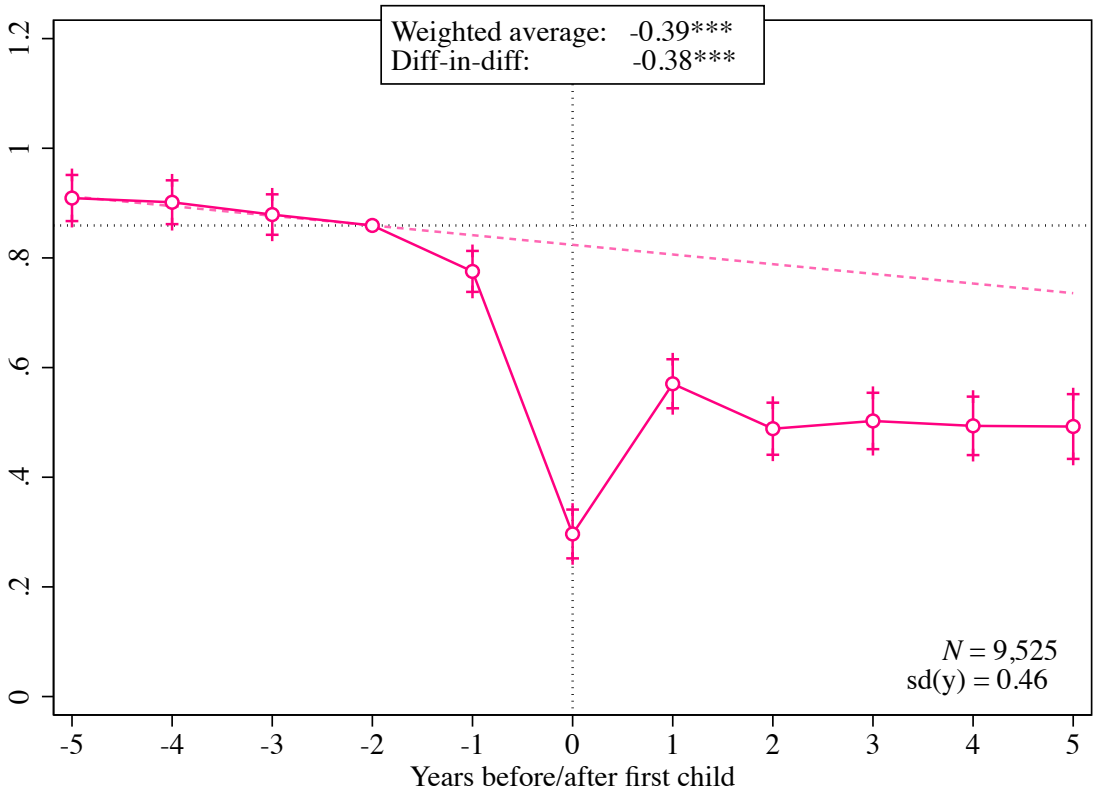
Source: OECD (`https://stats.oecd.org/Index.aspx?DataSetCode=LFS_SEXAGE_I_R#`). Data downloaded March 20th, 2018.

Figure 3.2: Share of young adults with tertiary education, by gender



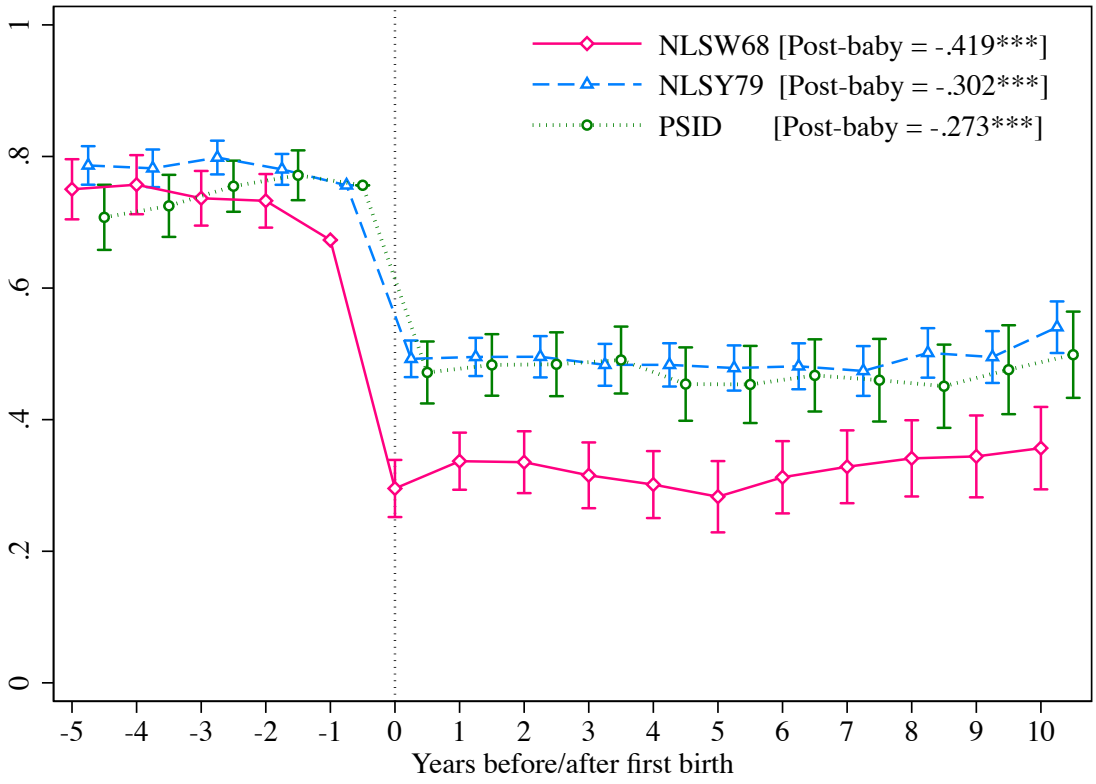
Source: CPS and EU-LFS.
Notes: The sample is restricted to individuals age 25 to 34. In the EU-LFS, tertiary education is defined as having completed ISCED levels 5 or 6. In the CPS, tertiary-education is defined as an associate's degree or more. In 1992, the CPS changed its coding of education. To account for this change, we set the share of people with tertiary education in 1991 to be the same as the share in 1992. For years prior to 1991, we set the share of people with tertiary education to be the share in 1992 times the percent change between that year and 1991.

Figure 3.3: Event-study analysis of how probability of employment changes after motherhood, UK



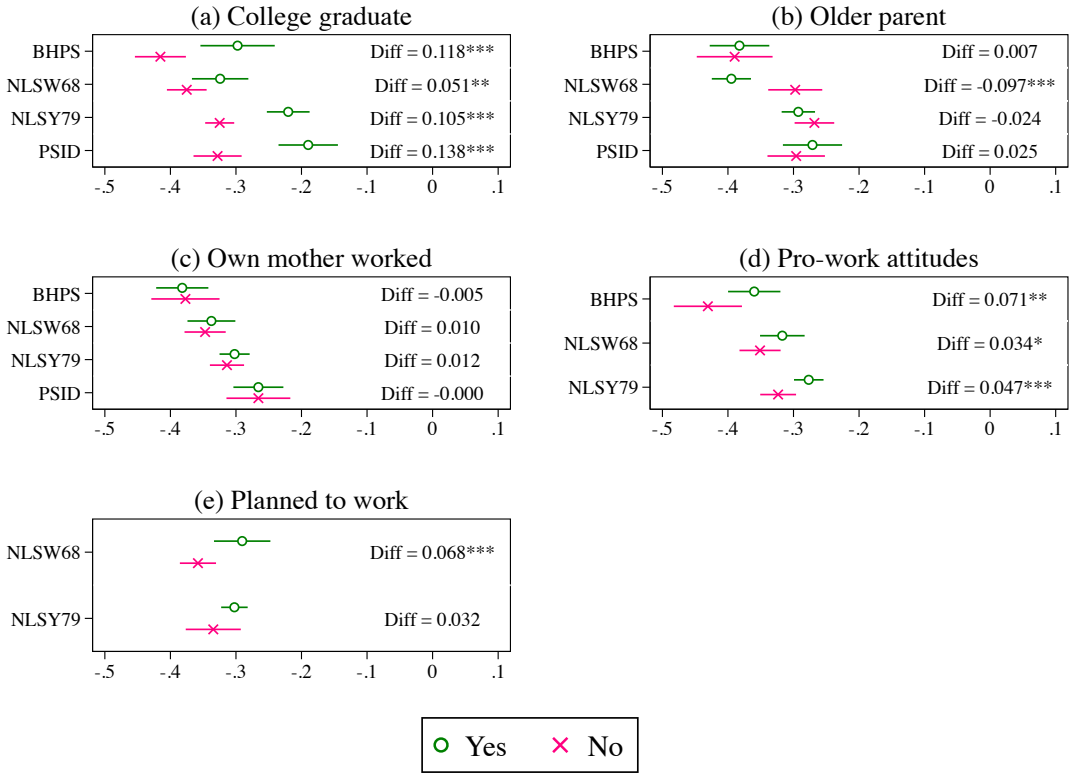
Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1.
Notes: The dependent dummy variable is a variable equal to 1 if an individual did paid work (i.e. was physically present at their job) last week. Event years are measured with some rounding error, so that event year 0 (the year of birth) is ±6 months around the first birth, so the child is either born or in utero. The year before birth is normalized to the mean of the outcome variable at event time $t = -1$ and $t = -5$ ($t = 5$) include all person-year observations five or more years before (five or more years after) the first birth. The event-study specification controls for age-in-years, calendar-year, and person fixed effects. Standard errors are clustered at the person level. The estimate reported at the top of the figure corresponds to a parallel difference-in-difference analysis where the event-time is collapsed into a simple before and after the birth of the child. The sample comprises women who were original sample members in the British Household Panel Survey and for whom we observe their first birth.
$^{***}p < 0.01;^{**}\ p < 0.05;^{*}\ p < 0.1$

Figure 3.4: Event-study analysis of how probability of employment changes after motherhood, US



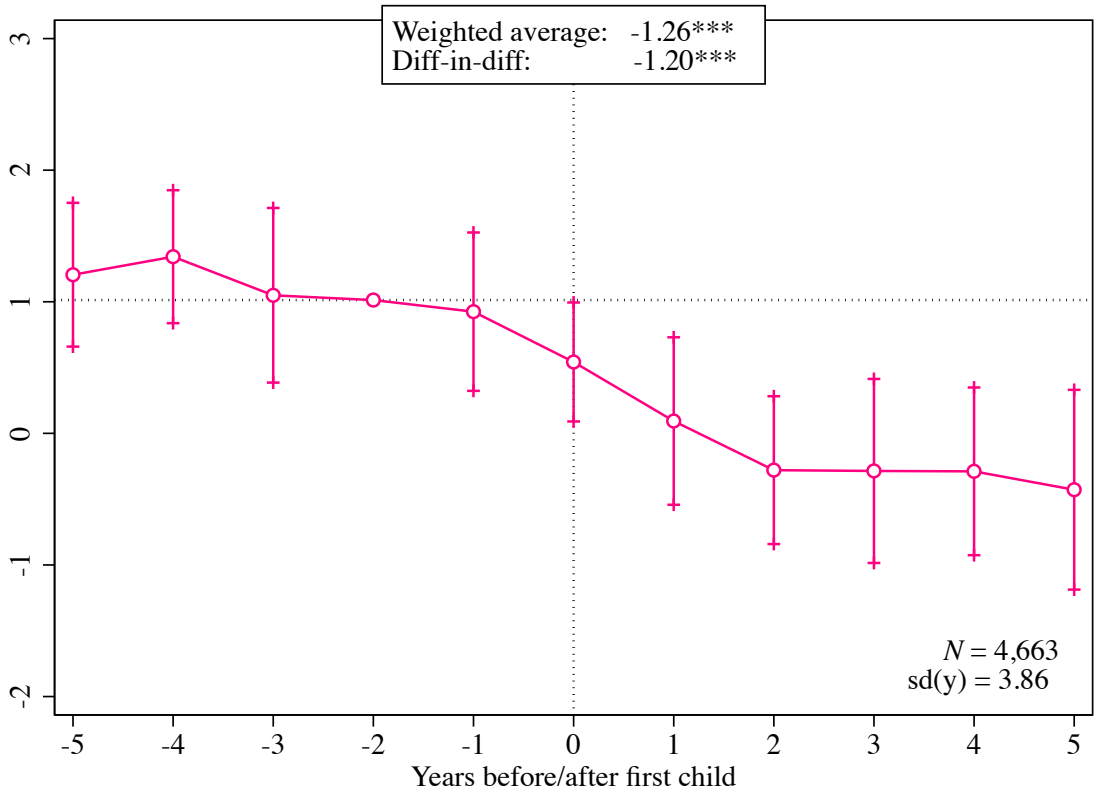
Sources: National Longitudinal Survey of Young Women (NLSW68), National Longitudinal Survey of Youth (NLSY79), and Panel Study of Income Dynamics (PSID). For more detail see Section 3.2.2 and Appendix A.2.2-A.2.4.
Notes: Sample comprises all female original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. The year before birth is normalized to the mean of the outcome variable at event time $t = -1$ and $t = -5$ ($t = 10$) include all person-year observations five or more years before (ten or more years after) the first birth. The event-study specification controls for age-in-years, calendar-year, and person fixed effects. The estimate reported at the top of the figure corresponds to a parallel difference-in-difference analysis where the event-time is collapsed into a simple before and after the birth of the child. Regressions are weighted by the probability of selection. Standard errors are clustered at the person level.
$^{***}p < 0.01;^{**}\ p < 0.05;^{*}\ p < 0.1$

Figure 3.5: Heterogeneous effects of first birth on current employment, women



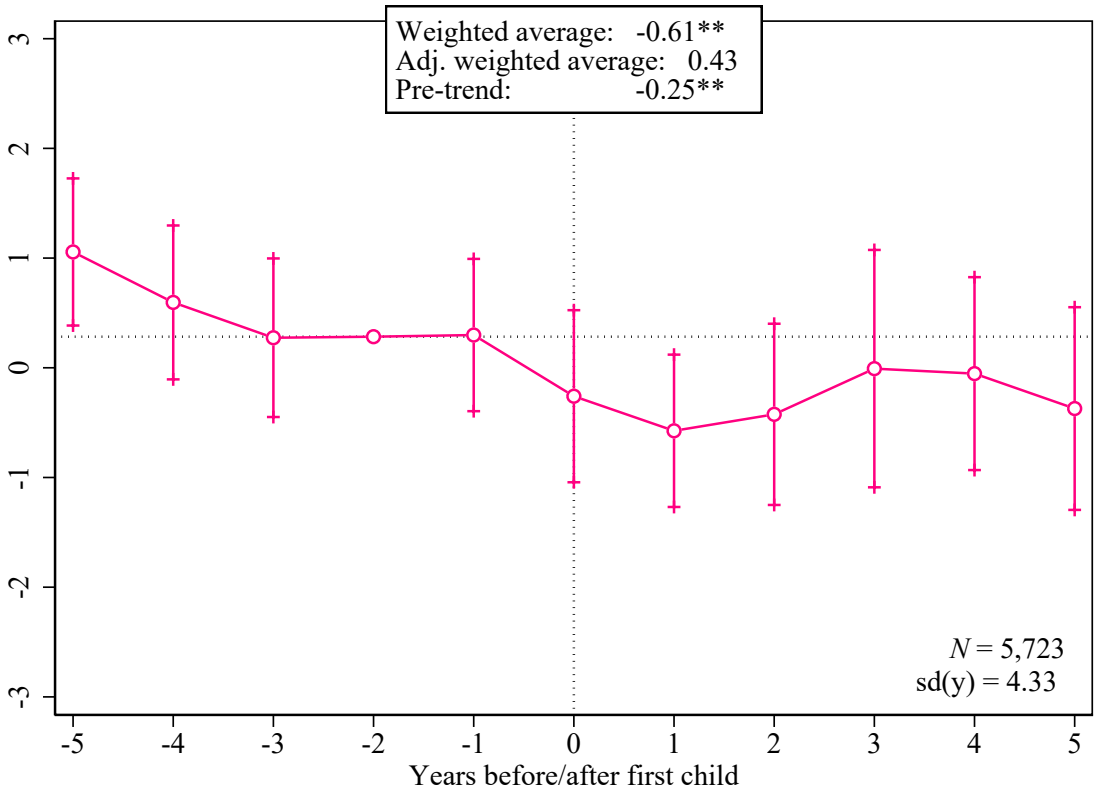
Sources: British Household Panel Survey (BHPS), National Longitudinal Survey of Young Women (NLSW68), National Longitudinal Survey of Youth (NLSY79), and Panel Study of Income Dynamics (PSID). For more detail see Section 3.2.2 and Appendix A.2.2-A.2.4.
Notes: Sample comprises all female original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Estimates are from a regression that is identical to our difference-in-differences specification (i.e., including person, age, and calendar year fixed effects), except that we also include an interaction term of *Post-baby* with our various heterogeneity variables (e.g. completed tertiary education). Estimates corresponding to "No" are from the *Post-baby* term and estimates corresponding to "Yes" are from the sum of the *Post* and interaction term. The "Diff" values listed to the right of each series of estimates are the difference between the "Yes" and "No" terms. Regressions in the NLSW68, NLSY79, and PSID are weighted by the probability of selection. Standard errors are clustered at the person level.

Figure 3.6: Event-study analysis of how gender-role attitudes change after motherhood (BHPS)



Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1. *Notes*: The gender-role index variable is constructed as the (standardized) sum of six questions on the appropriate role of men and women with respect to household, child-care, and labor-market responsibilities. It is increasing in the "pro-female-employment" direction. A list of these questions can be found in Appendix Table A.6 and event studies for each component can be found in Appendix Figure A.6. The sample comprises all female original sample members for whom we observe the event of childbirth. The year before birth is normalized to the mean of the outcome variable at event time $t = -1$ and $t = -5$ ($t = 5$) include all person-year observations five or more years before (five or more years after) the first birth. The event-study specification controls for age-in-years, calendar-year, and person fixed effects. The estimate reported at the top of the figure corresponds to a parallel difference-in-difference analysis where the event-time is collapsed into a simple before and after the birth of the child. Standard errors are clustered at the person level. $^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.1$

Figure 3.7: Event-study analysis of how gender-role attitudes change after motherhood (NLSY79)

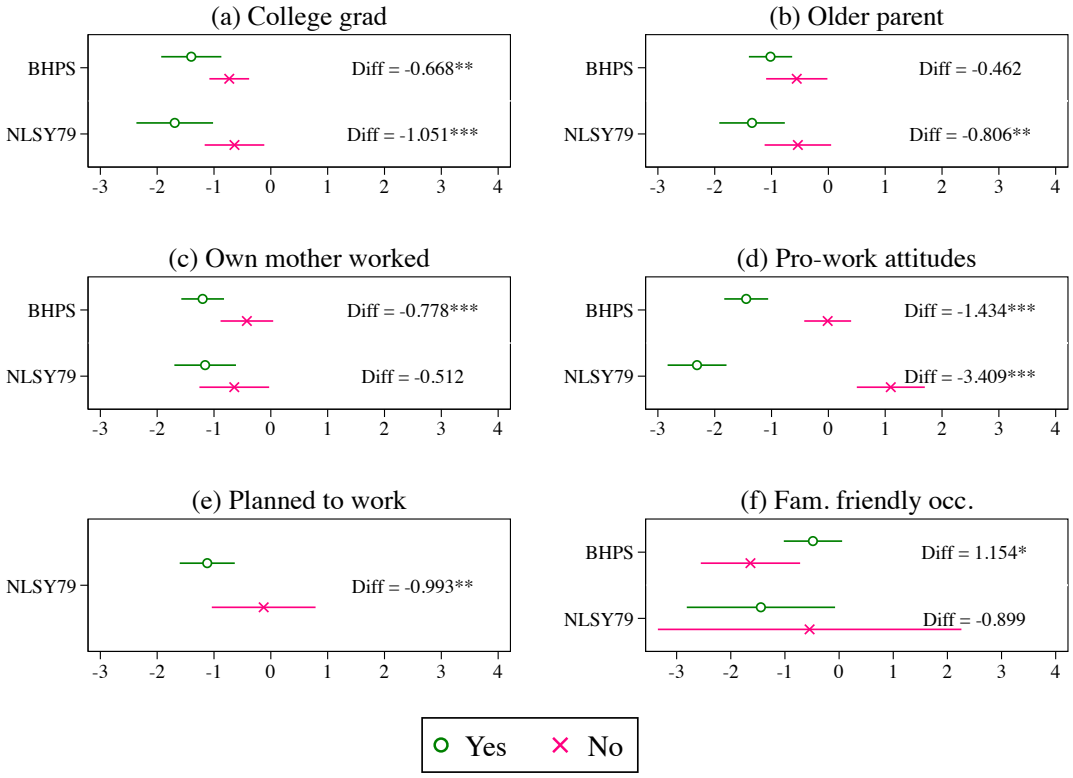


Source: National Longitudinal Survey of Youth. For more detail see Section 3.2.2 and Appendix A.2.3.

Notes: The gender-role index is constructed as the (standardized) sum of seven questions on the appropriate role of men and women with respect to household, child-care, and labor-market responsibilities. It is increasing in the "pro-female-employment" direction. A list of these questions can be found in Appendix Table A.8. The sample comprises all female original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. For a more detailed description of sampling criteria see Appendix A.2. The year before birth is normalized to the mean of the outcome variable at event time $t = -1$ and $t = -5$ ($t = 5$) include all person-year observations five or more years before (five or more years after) the first birth. The event-study specification controls for age-in-years, calendar-year, and person fixed effects. The estimate reported at the top of the figure corresponds to a parallel difference-in-difference analysis where the event-time is collapsed into a simple before and after the birth of the child. Standard errors are clustered at the person level.

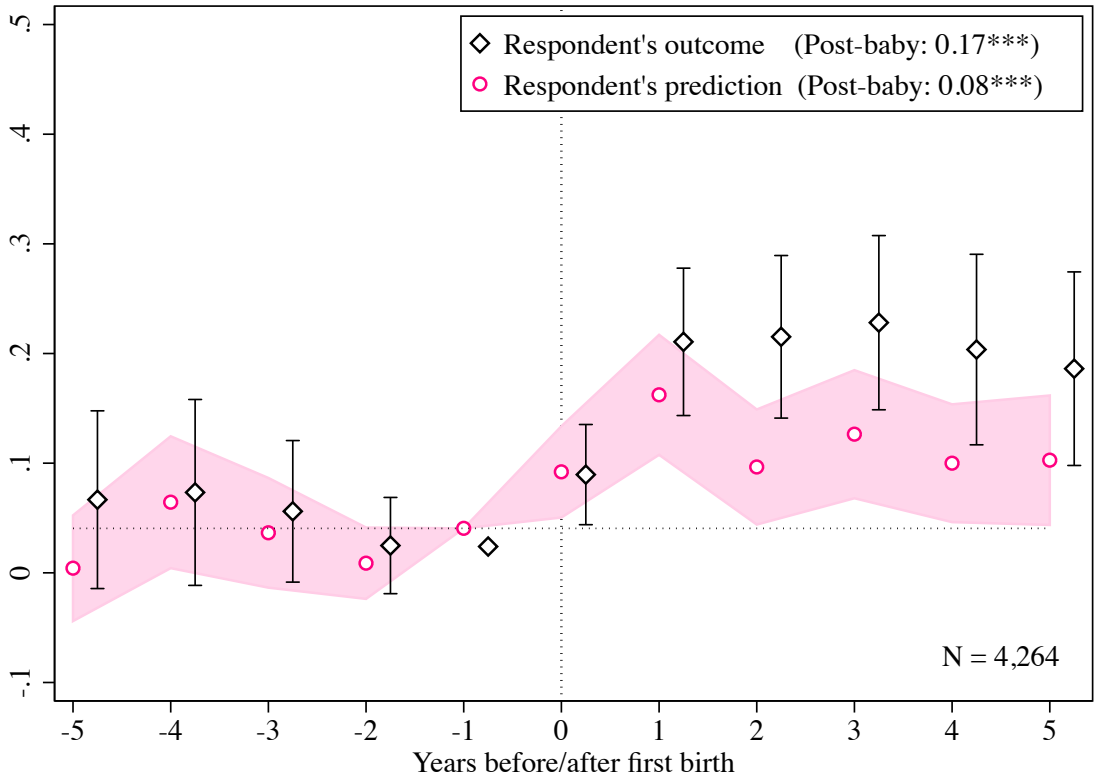
$^{***}p < 0.01;^{**}\ p < 0.05;^{*}\ p < 0.1$

Figure 3.8: Heterogeneous effects of first birth on women's gender-role attitudes



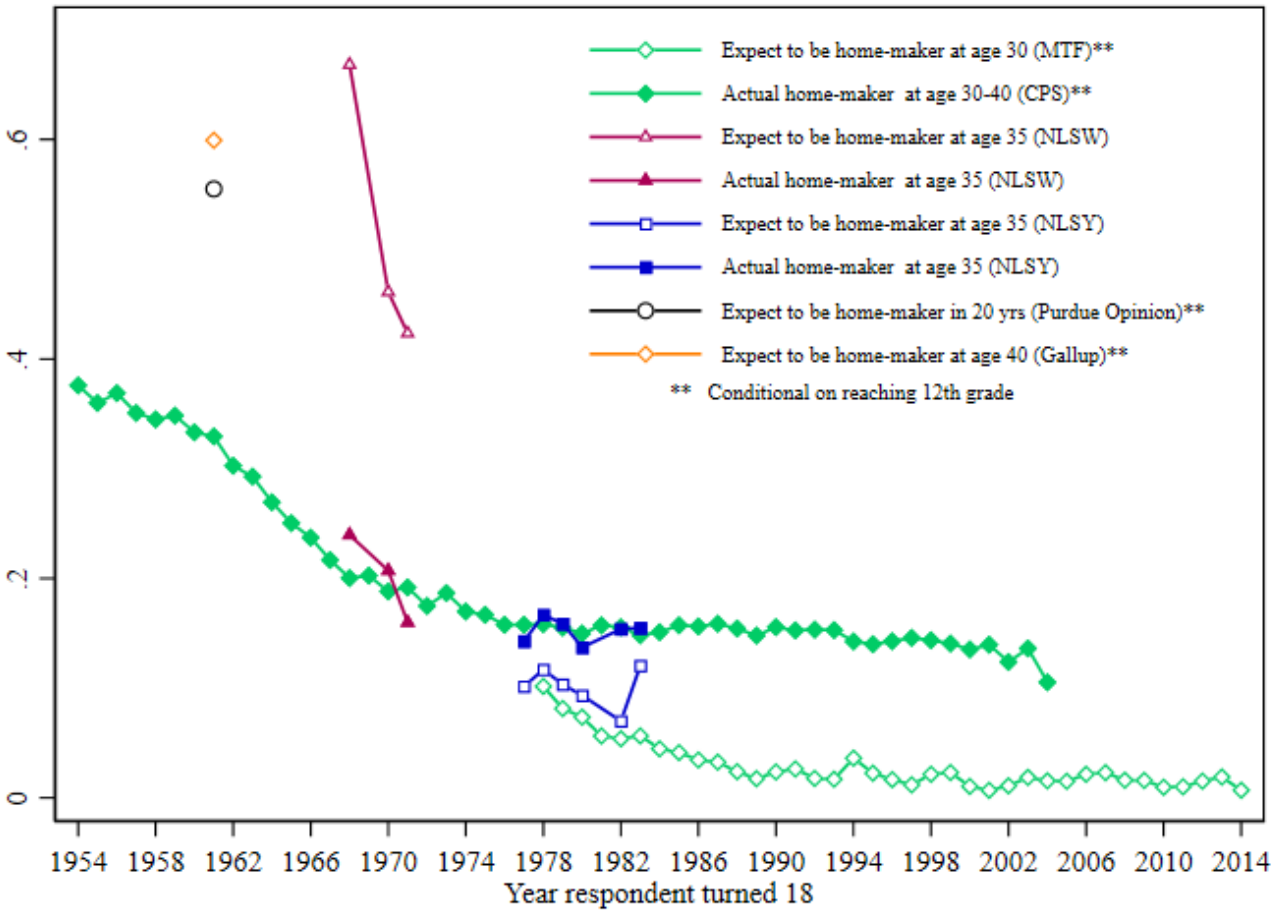
Sources: British Household Panel Survey (BHPS)and National Longitudinal Survey of Youth (NLSY79). For more detail see Section 3.2 and Appendix A.2.
Notes: Sample comprises all female original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Estimates are from a regression identical to our difference-in-differences specification (i.e., including person, age, and calendar-year fixed effects), except that we also include an interaction term of *Post-baby* with our various heterogeneity variables (e.g., obtained a college degree). Estimates corresponding to "No" are from the estimated coefficient on the *Post-baby* term and estimates corresponding to "Yes" are from the sum of the *Post-baby* and interaction term. The "Diff" values listed to the right of each series of estimates are the difference between the "Yes" and "No" terms. Regressions in the NLSY79 are weighted by the probability of selection. Standard errors are clustered at the person level and reported in brackets.

Figure 3.9: Event-study analysis of how expectations and reality of giving up paid work change upon motherhood, BHPS



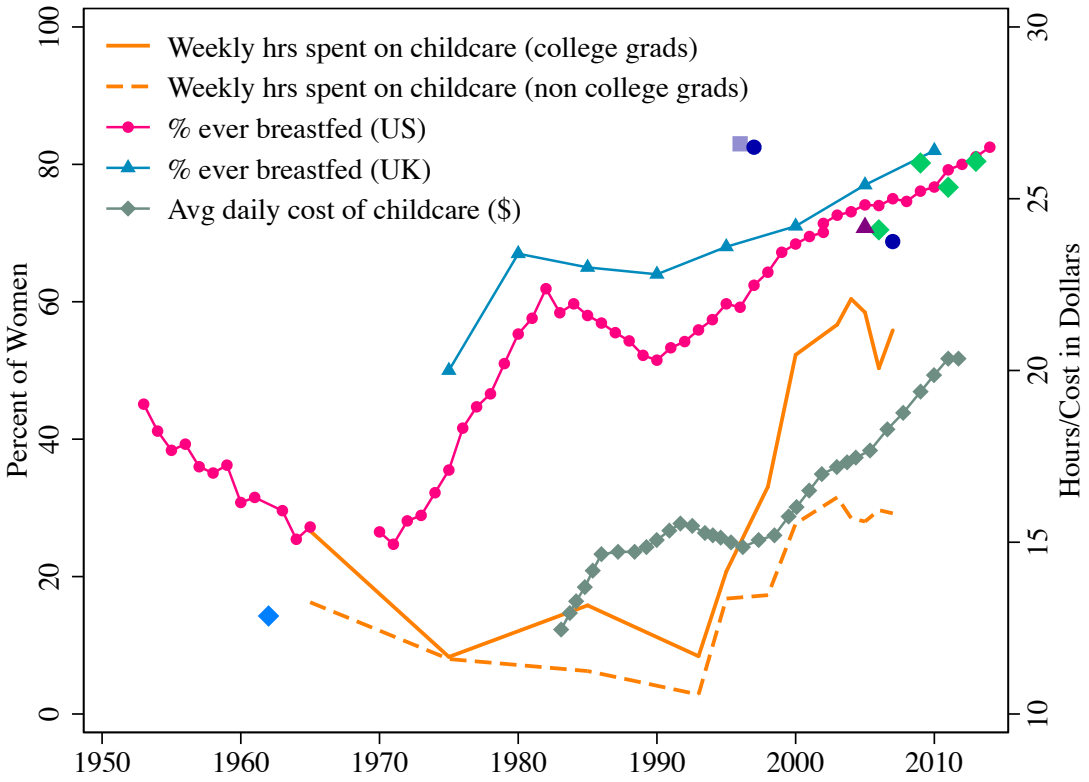
Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1. *Notes*: Sample includes only women who were observed working in the previous year. Predictions are made at $t = -1$ regarding labor supply in year t. Actual labor supply is whether a woman is working this year, conditional on having worked in the previous year. Sample comprises all female original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. The year before birth is normalized to the mean of the outcome variable at event time $t = -1$ and $t = -5$ ($t = 5$) include all person-year observations five or more years before (five or more years after) the first birth. The event-study specification controls for age-in-years, calendar-year, and person fixed effects. The estimate reported at the top of the figure corresponds to a parallel difference-in-difference analysis where the event-time is collapsed into a simple before and after the birth of the child. Standard errors are clustered at the person level.

Figure 3.10: Expectations and realizations of future home-making among U.S. high school seniors



Source: Monitoring the Future Survey (MTF), 1978 to 2014; CPS, 1972 to 2016; NLSW68, 1968 to 1973; NLSY79, 1979 to 1984; Purdue Opinion Panel, 1961; Gallup Opinion Poll, 1961.
Notes: Home-maker is defined in the CPS, NLSW68, and NLSY79 as an individual who is out of the labor force and married to a wage earner. In the MTF, Purdue Opinion Panel, and Gallup Opinion Poll, expected home-maker is defined as someone who reports planning to be a housewife or homemaker at age 30, 20 years from now, and at age 40, respectively. In the NLSW68, expected home-maker is someone who reports planning to be married and raising a family at age 35. In the NLSY79, expected home-maker is someone who reports planning to be married and raising a family *and* not working outside the home at age 35. To match the MTF, the CPS sample is restricted to individuals who have completed at least 11th grade. We similarly restrict our Purdue and Gallup samples.

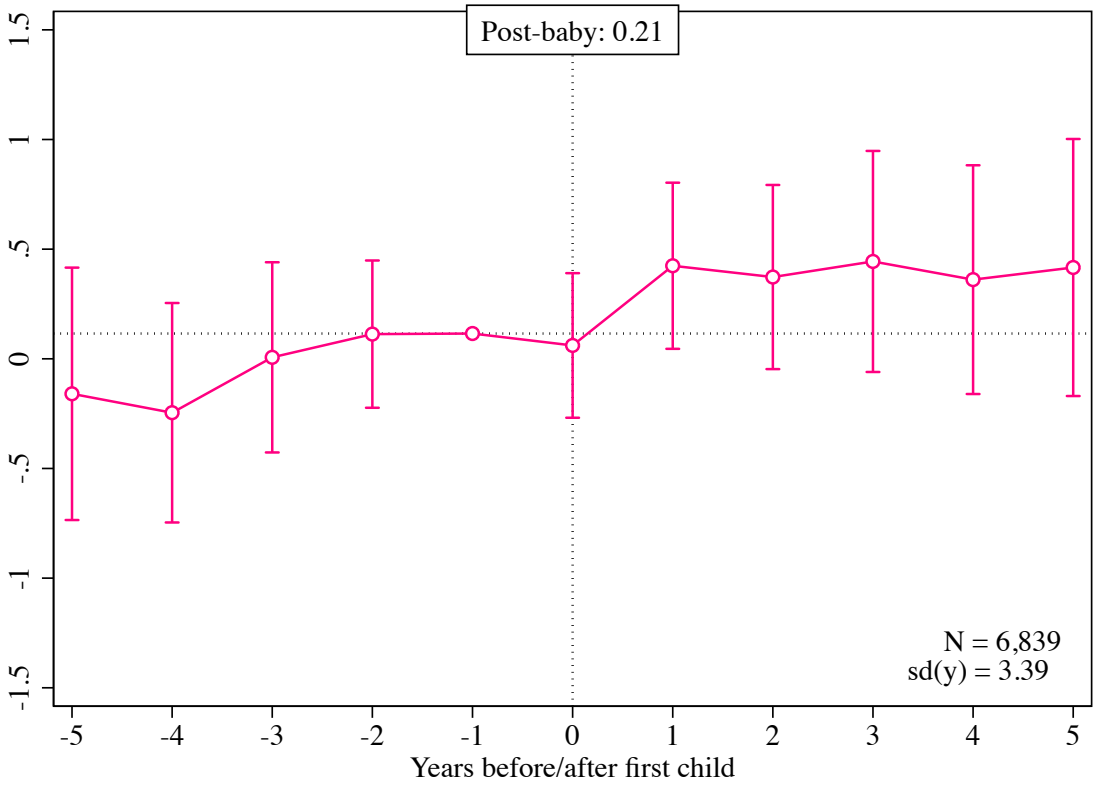
Figure 3.11: Various proxies for employment costs of motherhood at different child ages



Source: Breastfeeding data for the U.S. come from three different sources: the National Fertility Survey (1953-1965), the Ross Labs Mothers Survey (1970-2002), and the National Immunization Survey (2002-2014). Breastfeeding data for the UK come from the Infant Feeding Survey. Data for weekly hours spent on childcare are taken from Table 1 of (Ramey and Ramey, 2010). Data for average weekly childcare costs were extracted from Census Bureau Report CB13-62, available at https://www.census.gov/newsroom/press-releases/2013/cb13-62.html, using an online WebPlotDigitizer tool.

Notes: Estimates of breastfeeding rates from the National Fertility Survey, which was performed in 1965, 1970, and 1975, pool all three surveys and are calculated based on the year of birth of the child, rather than the year each woman was surveyed. We only plot years in which estimates come from 900 or more observations. In the Ross Labs Mothers Survey, the National Immunization Survey, and the Infant Feeding Survey, mothers are asked about their children who are currently infants. In these surveys, the survey year, rather than the year of birth of the child, is used. Estimates for weekly hours spent on childcare, taken from Ramey and Ramey (2010), come from a combination of twelve nationally representative time use surveys. As Ramey and Ramey (2010) document, much of the increase in the hours spent on childcare is driven by an increase in time spent on care of older and mixed age children, for example chauffeuring children to various activities. Cost of childcare measures the average weekly expenditures among people making childcare payments, in 2011 dollars.

Figure 3.12: Event-study analysis of how overall job satisfaction changes upon motherhood (BHPS)



Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1.
Notes: Higher values means greater overall job satisfaction. Components of the omnibus measure are: Overall job satisfaction as well as satisfaction with total pay, security, work itself, and hours worked. Sample comprises all female original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. The year before birth is normalized to the mean of the outcome variable at event time $t = -1$ and $t = -5$ ($t = 5$) include all person-year observations five or more years before (five or more years after) the first birth. The event-study specification controls for age-in-years, calendar-year, and person fixed effects. The estimate reported at the top of the figure corresponds to a parallel difference-in-difference analysis where the event-time is collapsed into a simple before and after the birth of the child. Standard errors are clustered at the person level.
$^{***}p < 0.01;^{**}\ p < 0.05;^{*}\ p < 0.1$

Table 3.1: Summary statistics for main panel data sources

	BHPS	NLSW 68	NLSY 79	PSID
–Employment				
Working this week	0.49	0.59	0.65	0.54
In work force this week	0.52	0.70	.	0.61
Hours worked last week	16.50	21.15	.	.
Expected to give up paid work last wave	0.05	.	.	.
Actually gave up paid work this wave	0.11	.	.	.
–Background				
Year of birth	1953.05	1948.28	1960.32	1969.59
Age at first birth	26.51	22.52	.	.
Total number of children	2.26	2.45	2.27	2.45
Married pre-baby	0.49	.	.	0.35
Black	.	0.28	0.29	0.35
College graduate	0.11	0.21	0.18	0.20
Planned to be working at 35	.	.	0.84	.
Gender-liberal pre-baby	0.48	0.39	0.57	.
Mom worked when resp. was 14	0.47	0.39	0.53	0.66
–Average # of Pre-/Post- Obs. per Person				
Number of employment observations, pre-baby	4.91	10.32	15.55	1.95
Number of employment observations, post-baby	0.32	1.76	0.90	1.17
Number of respondents	12379.00	5158.00	6283.00	39012.00
Observations	165,560	113,476	232,471	1521468

Source: British Household Panel Survey (BHPS), National Longitudinal Survey of Young Women (NLSW68), National Longitudinal Survey of Youth 1979 (NLSY79), Panel Survey of Income Dynamics (PSID). For more detail see Section 3.2 and Appendix A.2.1-A.2.4.
Notes: Sample comprises all original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Statistics in the NLSW68, NLSY79, and PSID are weighted by the probability of selection.

Table 3.2: Relationship between college education and probability of reporting that being a parent is harder than expected

	Dependent variable: Parenthood harder than expected				
	Women			Men	
	(1)	(2)	(3)	(4)	(5)
College Grad	0.118***	0.109***	0.0450**	0.0642	0.0251
	[0.0271]	[0.0304]	[0.0192]	[0.0507]	[0.0271]
Mean, dep. var.	0.520	0.520	0.520	0.377	0.414
Observations	1,638	1,634	5,983	554	2,448
Controls?	No	Yes	Yes	Yes	Yes
Age of children at home	0-5	0-5	0-17	0-5	0-17

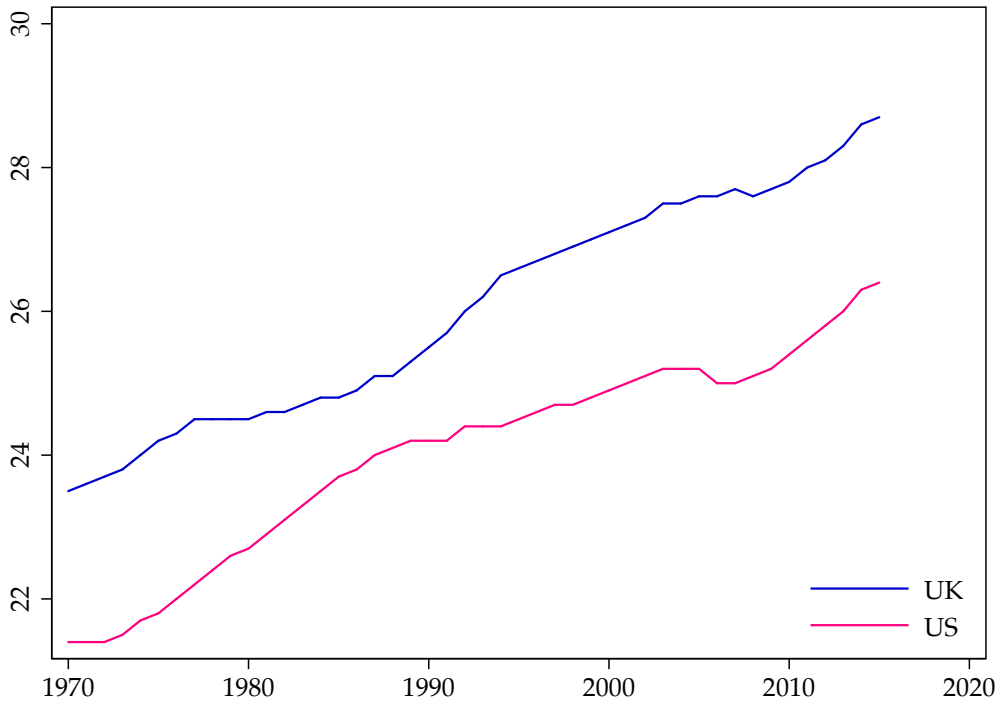
Source: Childhood Development Supplement to the Panel Study of Income Dynamics. For more detail see Appendix A.2.4.
Notes: Sample comprises all primary and secondary caregivers in the Childhood Development Supplement to the Panel Study of Income Dynamics. Columns (2)-(5) control for birth decile, survey year, race, child age, and number of children fixed effects, marital status fixed effects and log of household income. Columns (1), (2), and (4) restrict to parents with children under 6 whereas columns (3) and (5) restrict to parents with children under 18. Standard errors are reported in brackets.
$^{***}p < 0.01;^{**} p < 0.05;^{*} p < 0.1$

Appendix A

Mommy Effect Appendix

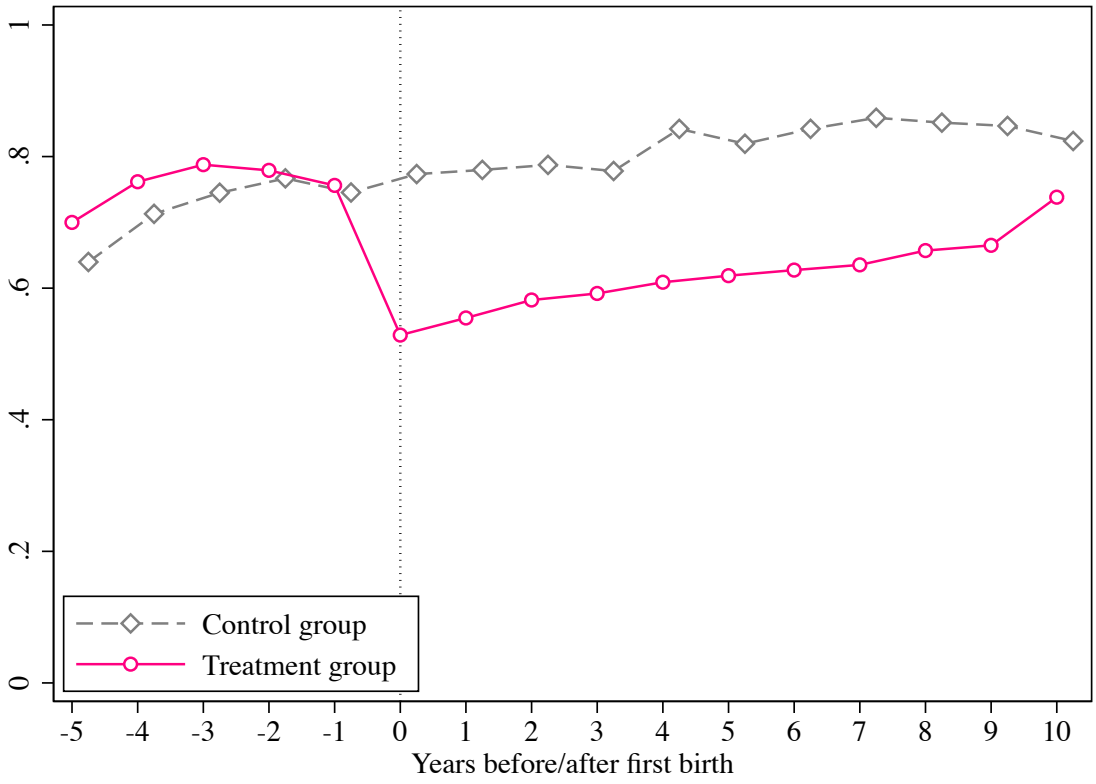
Appendix A.1. Appendix Figures and Tables

Appendix Figure A.1: Average age at first birth (U.S. and U.K. women)



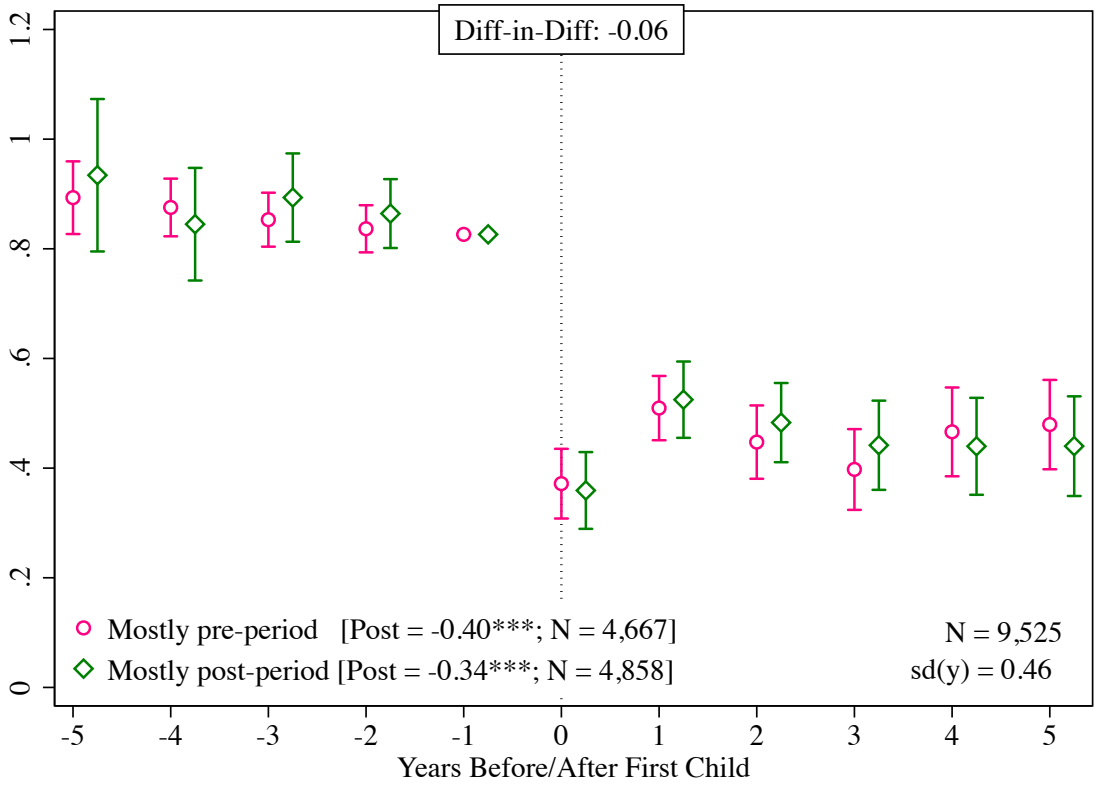
Source: Center for Disease Control (US) and Office for National Statistics (UK).

Appendix Figure A.2: Employment over event-time for women with and without children (NLSY), assigning "placebo births" to women without children



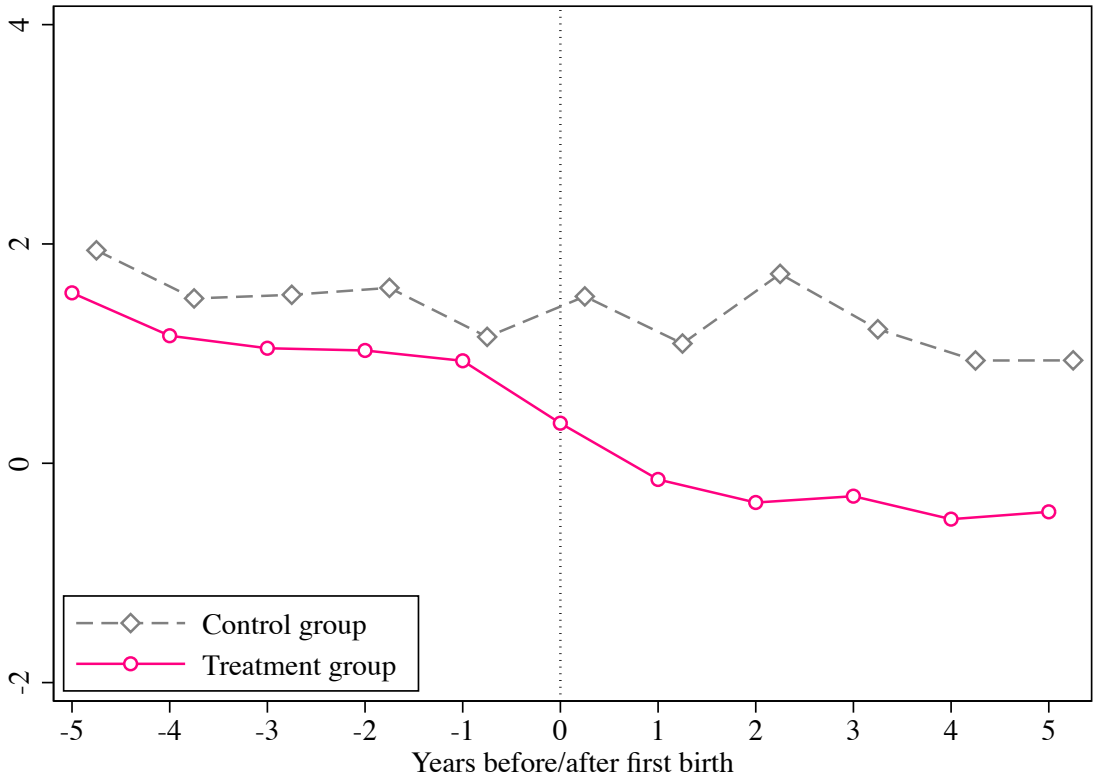
Source: NLSY.
Notes: For both series, raw means of the share currently employed are plotted. For women with children, event-time is defined relative to the birth of the first child. For women without children, we take the sample of women for whom we observed completed fertility (ages 20 to 40). We then randomly assign "placebo" dates of first birth based on what we observe in sample of mothers in the NLSY. In particular, we assume that women's age at first birth follows a log-normal distribution, where the mean and variance of this distribution are the observed mean and variance in age at first birth among NLSY mothers in our preferred analysis sample. We impose our standard sample restrictions to this control group (observed at least once pre- and post-birth, observed over ages 20–40, etc.) and then plot raw means with respect to "placebo event time."

Appendix Figure A.3: Employment event-study analysis, splitting respondents by whether they mostly identify the pre- or post-periods



Source: British Household Panel Survey.
Notes: The above graph plots estimates from an event study regression, which is our standard event study regression but also includes event study dummy variables interacted with a "mostly pre-period" variable. We construct our "mostly pre-period" variable by first calculating the share of total observations that are in the pre-period for each respondent. Then, we count someone as mostly identifying pre-period event-time coefficients if her share of pre-period observations is above the median across our preferred sample of women. In the BHPS, that median is 33%; thus, if a respondent's pre-period observations are greater than 33% of her total observations, then she is counted as identifying mostly pre-period observations. In the graph, the "mostly pre-period" series plots the event study coefficients plus the interacted (event-time) $\times$ (mostly pre-period) coefficients. The "mostly post-period" series plots just the standard event study coefficients. As with our standard event study regression, we include person, calendar-year, and age-in-years fixed effects and cluster our standard errors at the individual level.

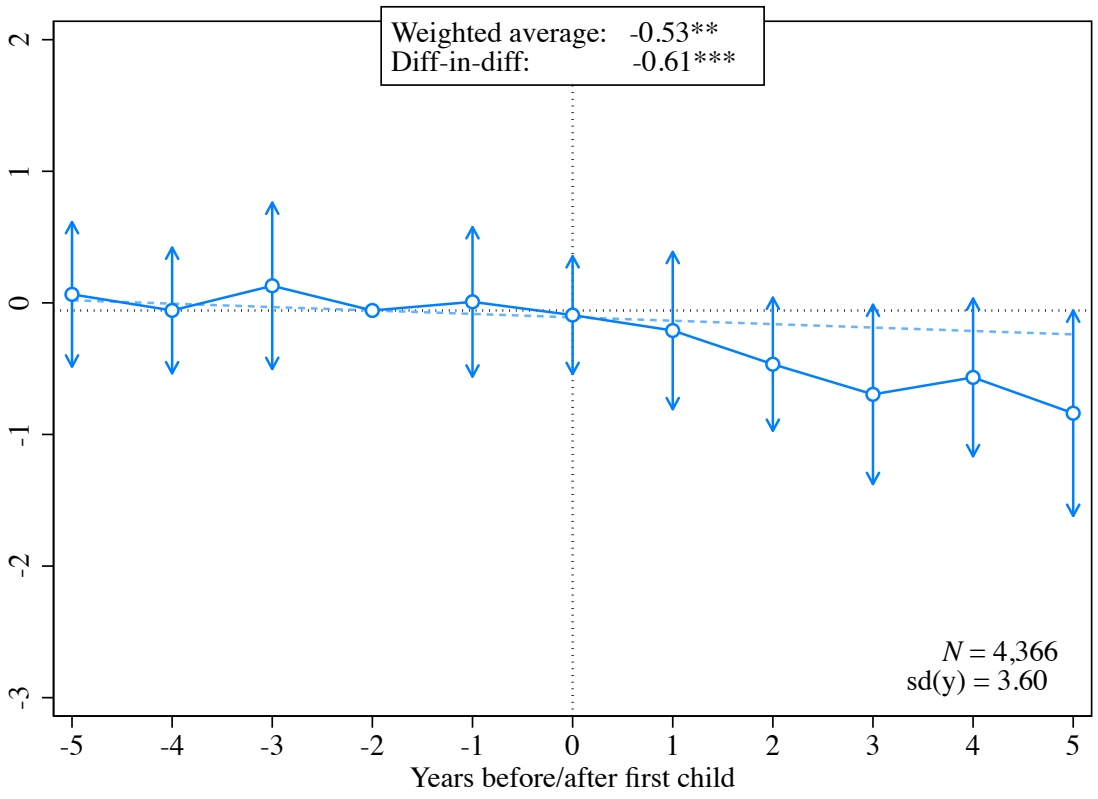
Appendix Figure A.4: Gender-norms index over event-time for women with and without children (BHPS), assigning "placebo births" to women without children



Source: BHPS.

Notes: For both series, raw means of the gender-roles index are plotted. For women with children, event time is defined relative to the birth of the first child. For women without children, we assign placebo births in the following manner (which differs somewhat from that in Figure A.2). Unlike the U.S. panel datasets, which extend over longer peiods, the BHPS is too short for us to examine women over the entire period of their key childbearing years, ages 20–40. So, unless we look only at older women, we cannot be sure that someone who is childless at the end of the panel in 2009 will not eventually have a child. We thus modify the procedure used in Appendix Figure A.2 and instead broadly follow Kleven *et al.* (2018). Our placebo group is thus defined as a group of women who are highly likely to remain childless (even if they are less than age forty in 2009). We generate these predictions by estimating a linear probability model of childlessness on a sample of women who have completed their fertility (i.e. are over age 40 at the end of our panel in 2009) and use estimates from this model to predict lifetime childlessness for women who have not yet completed their fertility (are under 40 in 2009). In particular, we regress childlessness on whether own mother worked, education-category fixed effects, income-quartile fixed effects, and birth-cohort fixed effects. We then select the top X% of women with respect to predicted childlessness and code them as being childless. We set X to equal the share of women in our data who are age forty and over and are childless (about 25 percent in our sample). We then assign this group random placebo first births in the same manner as in Appendix Figure A.2. We impose our standard sample restrictions to this control group (observed at least once pre- and post-birth, do not age into the sample, etc.) and then plot raw means with respect to "placebo event time."

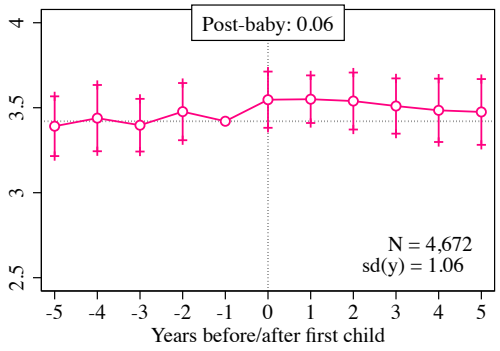
Appendix Figure A.5: Event-study analysis of how gender-role attitudes change after fatherhood (men in the BHPS)



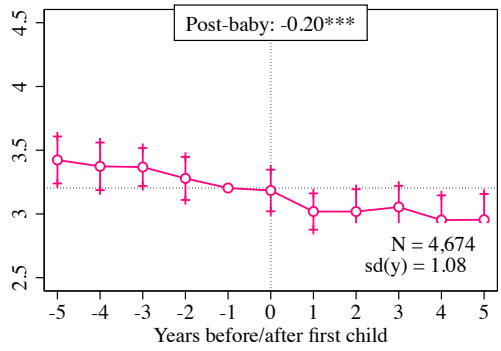
Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1. *Notes*: The gender-roles index is constructed as the (standardized) sum of six questions on the appropriate role of men and women with respect to child-care and labor-market responsibilities. It is increasing in the "pro-female-employment" direction. A list of these questions can be found in Appendix Table A.6. Sample comprises all male original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. The year before birth is normalized to the mean of the outcome variable at event time $t = -1$ and $t = -5$ ($t = 5$) include all person-year observations five or more years before (five or more years after) the first birth. The event-study specification controls for age-in-years, calendar-year, and person fixed effects. The estimate reported at the top of the figure corresponds to a parallel difference-in-difference analysis where the event-time is collapsed into a simple before and after the birth of the child. Standard errors are clustered at the person level.
$^{***}p < 0.01;^{**}\ p < 0.05;^{*}\ p < 0.1$

Appendix Figure A.6: Event-study analysis of how gender-role attitudes change after motherhood (if needed, responses are reversed so that all are increasing in the pro-female-employment position)

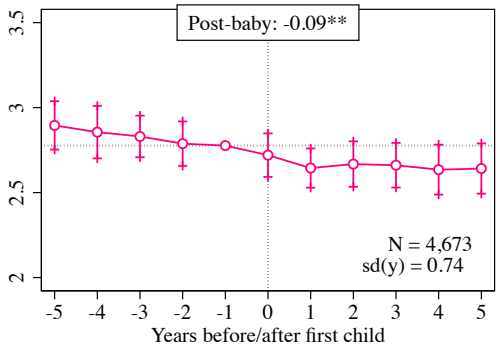
(a) Pre-school child suffers if mom works



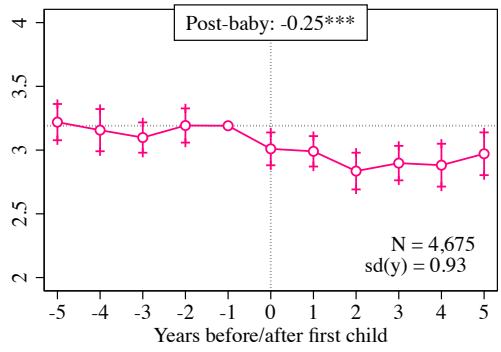
(b) Family suffers if mother works full-time



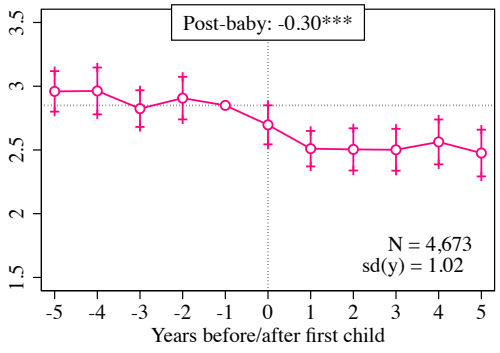
(c) Woman and family happier if she works



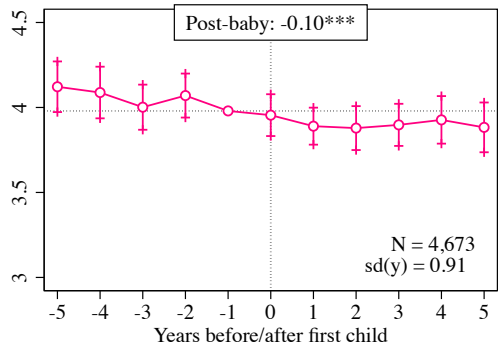
(d) Husband and wife should both contribute



(e) Full time job makes woman independent

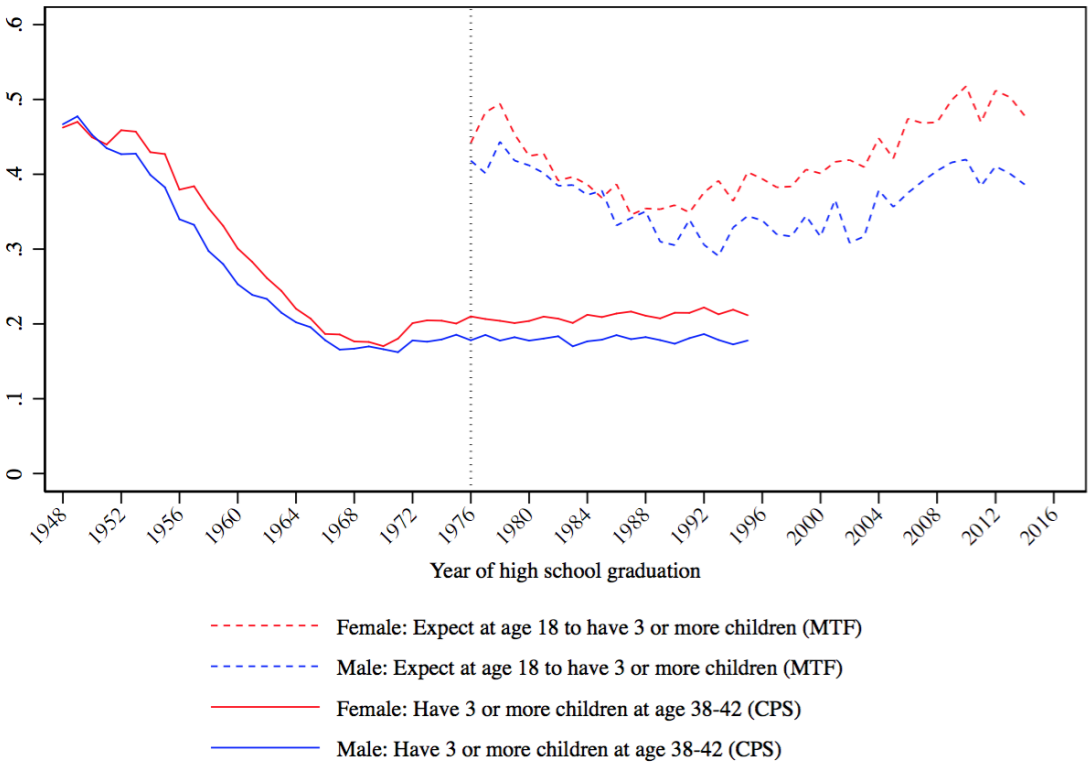


(f) Husband should earn, wife stay at home



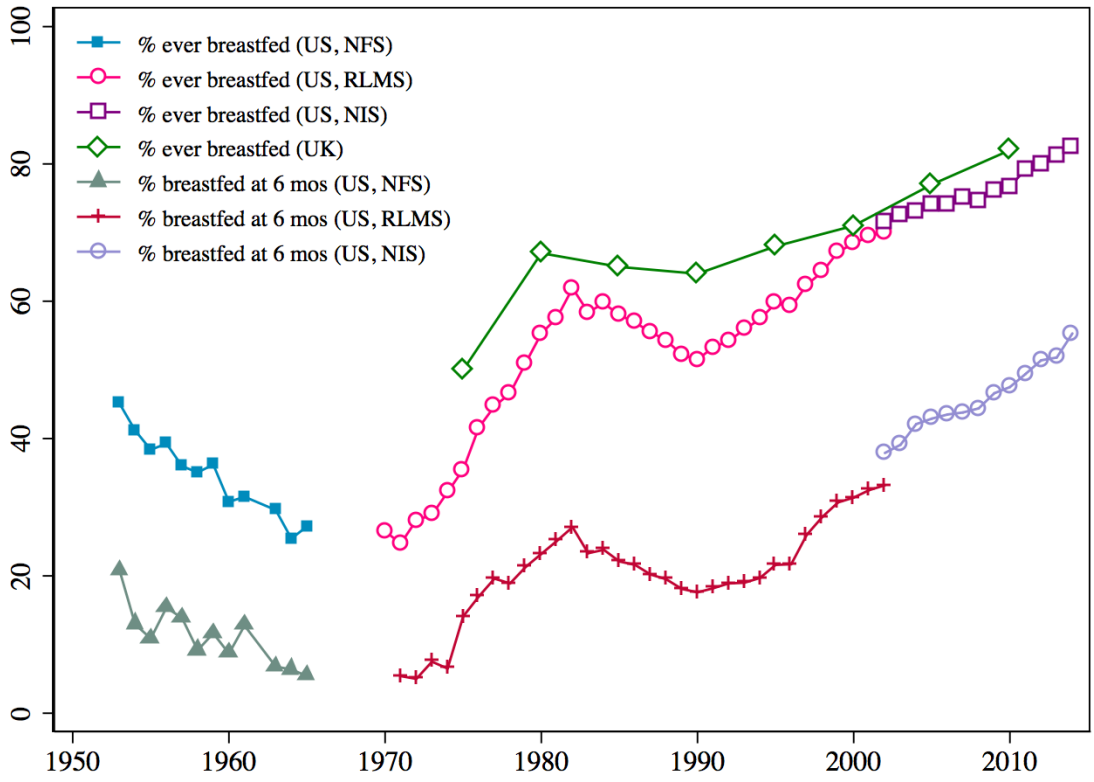
Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1. *Notes*: See notes to Figure 3.6. The analysis is identical here except instead of the standardized index of gender-role attitudes, we use as outcomes each component of the index. Unlike the omnibus measure, here we do not standardize the components to have mean zero and variance one.

Appendix Figure A.7: Expectations and realizations of fertility (number of children) at age 38 to 42 among U.S. high school seniors



Source: Monitoring the Future Survey, 1976 to 2014; CPS, 1970 to 2017.
Notes: The fertility variable in the MTF is coded from responses to the following question: "All things considered, if you could have exactly the number of children you want, what number would you choose to have?" The responses in the MTF are compared to the realized fertility of the same cohort of high school seniors when they are between the ages of 38 to 42 in the CPS. To match the MTF, the CPS sample is restricted to individuals who have completed at least 11th grade. The variable used to infer completed fertility in the CPS includes biological children, step-children, and adopted children. The observations are weighted using the appropriate sample weights from each survey.

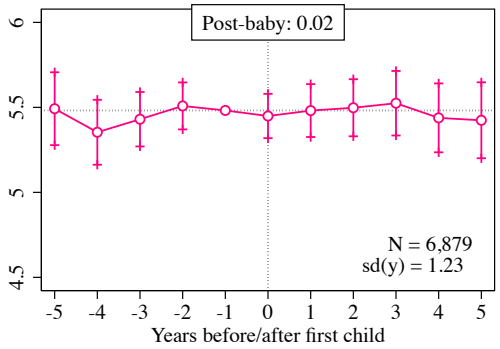
Appendix Figure A.8: Evolution of various measures of breastfeeding prevalence



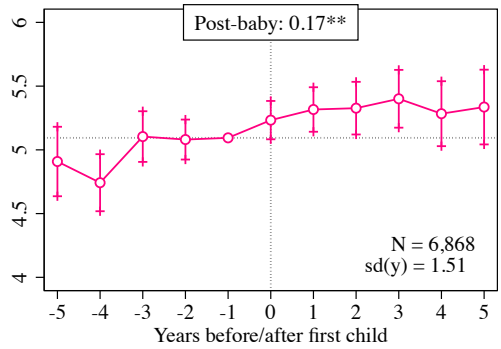
Source: Data sources for the US: Ross Labs Mothers Survey (RLMS), the National Immunization Survey (NIS), and the National Fertility Survey (NFS). Data source for the UK: Infant Feeding Survey (IFS).
Notes: Estimates of breastfeeding rates from the NFS, which was performed in 1965, 1970, and 1975, pool all three surveys and are calculated based on the year of birth of the child, rather than the year each woman was surveyed. We only plot years in which estimates come from 900 or more observations. In the RLMS, NIS, and IFS, mothers are asked about their children who are currently infants. In these surveys, the survey year, rather than the year of birth of the child, is used.

Appendix Figure A.9: Event-study analysis of how various factors of job satisfaction change after motherhood, BHPS (Higher values indicate greater satisfaction)

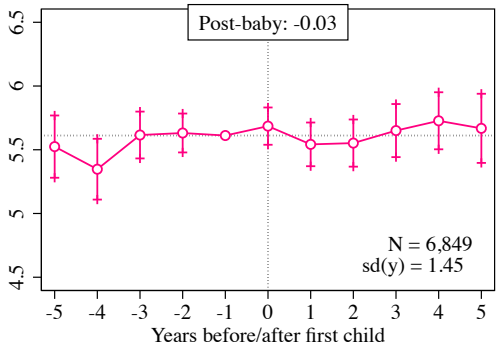
(a) Job satisfaction: Overall



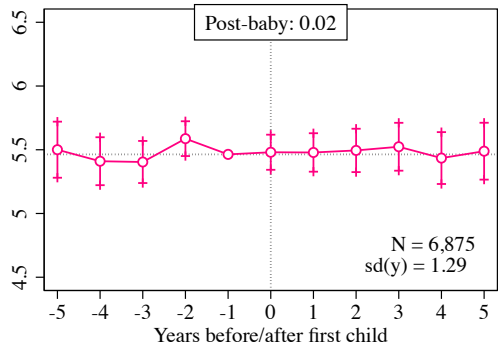
(b) Job satisfaction: Total pay



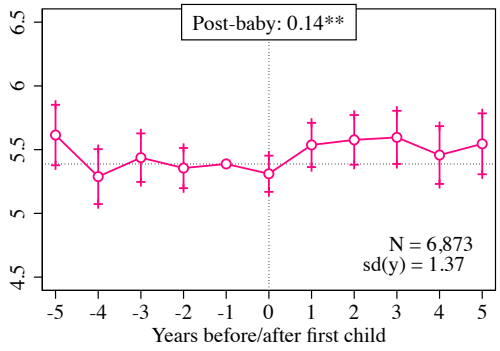
(c) Job satisfaction: Security



(d) Job satisfaction: Work itself



(e) Job satisfaction: Hours worked



Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1. *Notes:* The analysis is identical to that in Figure 3.12, except instead of the "overall" satisfaction variable, we examine questions related to satisfaction of specific attributes of the respondent's current job.

Appendix Table A.1: Summary statistics for main panel data sources, men

	BHPS	NLSY 79	PSID
–Employment			
Working this week	0.89	0.81	0.82
In work force this week	0.94	0.85	0.88
Hours worked last week	38.42	36.32	.
Expected to give up paid work last wave	0.01	.	.
Actually gave up paid work this wave	0.02	.	.
–Background			
Year of birth	1967.29	1961.70	1966.25
Age at first birth	30.45	25.96	26.29
Total number of children	1.78	2.29	2.24
Married pre-baby	0.60	0.58	0.61
Black	.	0.15	0.15
College graduate	0.27	0.26	0.39
Planned to be working at 35	.	0.88	.
Gender-liberal pre-baby	0.51	0.38	.
Mom worked when resp. was 14	0.67	0.54	0.58
–Average # of Pre-/Post- Observations per Person			
Number of employment observations, pre-baby	4.52	9.48	8.96
Number of employment observations, post-baby	6.46	14.52	12.18
Number of respondents	628.00	2308.00	796.00
Observations	10,034	55,087	16,605

Source: British Household Panel Survey (BHPS), National Longitudinal Survey of Youth 1979 (NLSY79), Panel Survey of Income Dynamics (PSID). For more detail see Section 3.2 and Appendix A.2.1-A.2.4.

Notes: Sample comprises all male original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. For a more detailed description of sampling criteria for each dataset see Appendix A.2. Statistics for the NLSY79 and PSID are weighted by the probability of selection.

Appendix Table A.2: Effect of first birth on employment last week, robustness checks, BHPS and NLSW68

	Dependent variable: Employed last week								
	BHPS					NLSW68			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post-baby	-0.381***	-0.381***	-0.369***	-0.338***	-0.376***	-0.419***	-0.427***	-0.364***	-0.389***
	[0.0195]	[0.0195]	[0.0178]	[0.0176]	[0.0204]	[0.0158]	[0.0166]	[0.0127]	[0.0149]
Pre-baby mean, dep. var.	0.826	0.826	0.826	0.835	0.709	0.624	0.624	0.624	0.624
Observations	9,525	9,525	9,525	13,111	8,940	26,308	26,308	26,308	32,264
+ Indiv. lin. time trend	No	Yes	No	No	No	No	Yes	No	No
+ Drop person FE	No	No	Yes	No	No	No	No	Yes	No
+ Control group	No	No	No	Yes	No	No	No	No	Yes
+ Exclude age outliers	No	No	No	No	Yes	n/a	n/a	n/a	n/a

Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1.
Notes: Sample comprises all original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Age-in-years, calendar year, and person fixed effects included in all regressions unless otherwise noted. "Exclude age outliers" in the BHPS drops women who have their first child before the age of 20 or after the age of 40. Regressions for NLSW68 are weighted by the probability of selection; regressions for BHPS are not weighted, but inclusion of weights to not substantively change estimated coefficient. (Results available on request.) Standard errors are clustered at the person level and reported in brackets.
$^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.10$

Appendix Table A.3: Effect of first birth on current employment, robustness checks, PSID and NLSY79

	Dependent variable: Employed last week								
	PSID					NLSY79			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post-baby	-0.227***	-0.224***	-0.266***	-0.227***	-0.260***	-0.277***	-0.276***	-0.314***	-0.277***
	[0.0153]	[0.0156]	[0.0138]	[0.0153]	[0.0191]	[0.00969]	[0.00973]	[0.00941]	[0.00969]
Pre-baby mean, dep. var.	0.543	0.543	0.543	0.543	0.570	0.675	0.675	0.675	0.675
Observations	22,458	22,458	22,458	22,458	15,802	54,175	54,175	54,175	54,175
+ Indiv. lin. time trend	No	Yes	No	No	No	No	Yes	No	No
+ Drop person FE	No	No	Yes	No	No	No	No	Yes	No
+ Control group	No	No	No	Yes	No	No	No	No	Yes
+ Agers-In	Yes	Yes	Yes	Yes	No	n/a	n/a	n/a	n/a

Source: National Longitudinal Survey of Young Women (NLSW68), National Longitudinal Survey of Youth 1979 (NLSY79), Panel Survey of Income Dynamics (PSID) For more detail see Section 3.2.2 and Appendix A.2.2-A.2.4.

Notes: Sample comprises all female original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. "Agers-in" refers to children of original sample households who become eligible for interviews as they come of age. Age-in-years, calendar year, and person fixed effects included in all regressions unless otherwise noted. Regressions are weighted by the probability of selection. Standard errors are clustered at the person level and reported in brackets.

$^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.10$

Appendix Table A.4: Effect of first birth on current employment, U.S. and U.K. men

	Dependent var: Employed last week		
	BHPS	NLSY79	PSID
	(1)	(2)	(3)
Post-baby	-0.0332***	0.0156*	0.00709
	[0.0126]	[0.00866]	[0.0146]
Pre-baby mean, dep. var.	0.860	0.659	0.586
Observations	8,980	55,087	13,241

Source: British Household Panel Survey (BHPS), National Longitudinal Survey of Youth 1979 (NLSY79), Panel Survey of Income Dynamics (PSID). For more detail see Section 3.2 and Appendix A.2.1-A.2.4.
Notes: Sample comprises all male original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Age-in-years, calendar-year, and person fixed effects included in all regressions unless otherwise noted. Regressions are weighted by the probability of selection. Standard errors are clustered at the person level and reported in brackets.
$^{***}p < 0.01;^{**}\ p < 0.05;^{*}\ p < 0.10$

Appendix Table A.5: Impacts of second baby on current employment, women

	Dependent variable: Employed last week							
	BHPS		NLSW68		NLSY79		PSID	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-baby	-0.373***	-0.365***	-0.308***	-0.317***	-0.287***	-0.285***	-0.258***	-0.266***
	[0.0193]	[0.0202]	[0.0262]	[0.0270]	[0.0117]	[0.0116]	[0.0201]	[0.0202]
2+ kids	-0.0986***	-0.141***	-0.126***	-0.126***	-0.0849***	-0.128***	-0.0645***	-0.118***
	[0.0207]	[0.0253]	[0.0263]	[0.0270]	[0.0115]	[0.0123]	[0.0194]	[0.0201]
Pre-baby mean, dep. var.	0.826	0.826	0.624	0.624	0.675	0.675	0.541	0.541
Observations	9,525	9,525	26,308	26,308	54,175	54,175	22,342	22,342
+ Indiv. lin. time trend	No	Yes	No	Yes	No	Yes	No	Yes

Sources: British Household Panel Survey (BHPS), National Longitudinal Survey of Young Women (NLSW68), National Longitudinal Survey of Youth (NLSY79), and Panel Study of Income Dynamics (PSID). For more detail see Section 3.2 and Appendix A.2.1-A.2.4.
Notes: Sample comprises all female original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Age-in-years, calendar-year, and person fixed effects included in all regressions unless otherwise noted. Regressions in the NLSW68, NLSY79, and PSID are weighted by the probability of selection. Standard errors are clustered at the person level and reported in brackets. $^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.1$

Appendix Table A.6: Summary statistics, gender-role attitudes (BHPS)

	Men	Women
–Gender Norms (Higher = More Liberal, Scale: 1-5)		
A pre-school child is likely to suffer if his or her mother works (Rev.)	3.07	3.39
All in all, family life suffers when the woman has a full time job (Rev.)	3.27	3.39
A woman and her family would all be happier if she goes out to work	2.85	2.88
Both the husband and wife should contribute to the household income	3.33	3.41
Having a full-time job is the best way for a woman to be an independent person	3.11	3.03
Husband's job is to earn money; wife's job is to look after home/family (Rev.)	3.78	4.06
Omnibus gender norms (higher = more liberal)	-0.48	0.29
–Average # of Pre-/Post- Observations per Person		
Number of omnibus gender observations, pre-baby	2.48	2.22
Number of omnibus gender observations, post-baby	2.80	3.01
Number of respondents	628.00	681.00

Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1.
Notes: Sample comprises all original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Individual component attitudes are on a scale of 1 to 5 (1 = strongly agree; 5 = strongly disagree) and reversed so that higher values imply more liberal, less traditional attitudes toward gender roles. The omnibus gender-role index takes each component question, standardizes it to have mean zero and standard deviation one, and sums across all questions.

Appendix Table A.7: Effect of motherhood on individual components of the gender-role index, controlling for a person-specific linear time trend (BHPS)

Dependent variable:	Child Suffers		Family Suffers		Family Happier	
	(1)	(2)	(3)	(4)	(5)	(6)
Post-baby	0.0697	0.0697	-0.199***	-0.200***	-0.0849*	-0.0850*
	[0.0474]	[0.0474]	[0.0480]	[0.0480]	[0.0348]	[0.0348]
Mean, dep. var.	3.391	3.391	3.393	3.393	2.877	2.877
Observations	4,720	4,720	4,722	4,722	4,721	4,721
+ Linear time trend	No	Yes	No	Yes	No	Yes

Dependent variable:	Both Contribute		Full-time Indep.		Husband Earn	
	(1)	(2)	(3)	(4)	(5)	(6)
Post-baby	-0.248***	-0.248***	-0.301***	-0.301***	-0.0914*	-0.0916*
	[0.0420]	[0.0420]	[0.0461]	[0.0461]	[0.0361]	[0.0361]
Mean, dep. var.	3.415	3.415	3.035	3.035	4.059	4.059
Observations	4,723	4,723	4,721	4,721	4,721	4,721
+ Linear time trend	No	Yes	No	Yes	No	Yes

Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1.
Notes: If needed, responses are reversed so that all are increasing in the pro-female-employment position but otherwise are not standardized. Sample comprises all original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Age-in-years, calendar-year, and person fixed effects included in all regressions unless otherwise noted. Standard errors are clustered at the person level and reported in brackets.
$^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.1$

Appendix Table A.8: Summary statistics, gender-role attitudes (NLSY79)

	Men	Women
–Gender Norms (Higher = More Liberal, Scale: 1-5)		
A woman's place is in the home, not in the office or shop	3.04	3.37
A wife who carries out her family responsibilities doesn't have time for a job	2.86	3.10
A working wife feels more useful than one who doesn't hold a job (Rev.)	2.74	2.51
The employment of wives leads to more juvenile delinquency	2.87	2.98
It's better if the man is the achiever, the woman takes care of the family	2.72	2.90
Men should share the work around the house with women (Rev.)	3.12	3.35
Women are happier if they stay home and take care of their children	2.75	2.92
Omnibus gender norms (higher = more liberal)	-0.46	0.96
–Average # of Pre-/Post- Observations per Person		
Number of omnibus gender observations, pre-baby	2.28	2.13
Number of omnibus gender observations, post-baby	1.51	1.69
Number of respondents	1303.00	1456.00

Source: National Longitudinal Survey of Youth, 1979. For more detail see Section 3.2.2 and Appendix A.2.3.
Notes: Sample comprises all original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Statistics are weighted by the probability of selection. Individual component attitudes are on a scale of 1 to 5 (1= strongly agree; 5 = strongly disagree) and reversed so that higher values imply more liberal, less traditional attitudes toward gender roles. The omnibus gender-roles index takes each component question, standardizes it to have mean zero and standard deviation one, and sums across all questions. Question wording is altered for readability, see A.2.3 for exact wording.

Appendix Table A.9: Effect of first and second birth on gender-role attitudes, UK

	Dependent variable: Omnibus gender-roles index							
	Women						Men	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-baby	-0.921***	-1.137***	-1.178***	-0.853***	-0.874***	-0.880***	-0.250	-0.174
	[0.158]	[0.236]	[0.225]	[0.149]	[0.158]	[0.226]	[0.158]	[0.157]
2+ kids					-0.460***	-0.832***		-0.735***
					[0.170]	[0.249]		[0.188]
Pre-baby mean, dep. var.	1.168	1.168	1.168	1.186	1.168	1.168	-0.0679	-0.0679
Observations	4,663	4,663	4,663	6,414	4,663	4,663	4,366	4,366
+ Indiv. lin. time trend	No	Yes	No	No	No	Yes	No	No
+ Drop person FE	No	No	Yes	No	No	No	No	No
+ Control Group	No	No	No	Yes	No	No	No	No

Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1.
Notes: Sample comprises all original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Age-in-years, calendar-year, and person fixed effects included in all regressions unless otherwise noted. Regressions are weighted by the probability of selection. Standard errors are clustered at the person level and reported in brackets.
$^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.1$

Appendix Table A.10: Effect of first birth and second birth on gender-role attitudes, US (NLSY79)

	Dependent variable: Omnibus gender-roles index							
	Women						Men	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-baby	-0.966***	-0.980***	-1.639***	-0.883***	-0.922***	-0.935***	-0.498**	-0.451*
	[0.244]	[0.247]	[0.234]	[0.191]	[0.249]	[0.251]	[0.251]	[0.255]
2+ kids					-0.198	-0.200		-0.214
					[0.278]	[0.278]		[0.285]
Pre-baby mean, dep. var.	0.881	0.881	0.881	0.845	0.881	0.881	-1.077	-1.077
Observations	5,493	5,493	5,493	8,545	5,493	5,493	4,834	4,834
+ Indiv. lin. time trend	No	Yes	No	No	No	Yes	No	No
+ Drop person FE	No	No	Yes	No	No	No	No	No
+ Control Group	No	No	No	Yes	No	No	No	No

Source: National Longitudinal Survey of Youth (NLSY79). For more detail see Section 3.2.2 and Appendix A.2.3.
Notes: Sample comprises all original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Age-in-years, calendar-year, and person fixed effects included in all regressions unless otherwise noted. Regressions are weighted by the probability of selection. Standard errors are clustered at the person level and reported in brackets.
$^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.1$

Appendix Table A.11: Expected to work this year but wrong, robustness checks, U.K. men and women

	Dependent variable: Expected to work but wrong							
	Women						Men	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-baby	0.160***	0.144***	0.148***	0.143***	0.161***	0.141***	-0.00471	-0.00508
	[0.0279]	[0.0181]	[0.0179]	[0.0256]	[0.0279]	[0.0189]	[0.0130]	[0.0129]
2+ kids					0.0314	0.00866		-0.00910
					[0.0297]	[0.0147]		[0.0134]
Pre-baby mean, dep. var.	0.0744	0.0744	0.0744	0.0678	0.0744	0.0744	0.0280	0.0280
Observations	3,785	3,785	3,785	5,314	3,785	3,785	4,122	4,122
+ Indiv. lin. time trend	No	Yes	No	No	No	Yes	No	No
+ Drop person FE	No	No	Yes	No	No	No	No	No
+ Control Group	No	No	No	Yes	No	No	No	No

Source: British Household Panel Survey. For more detail see Section 3.2.1 and Appendix A.2.1.
Notes: Sample comprises all original sample members for whom we observe the event of childbirth. Individuals must also be observed during prime child-bearing years. Age in years, calendar year, and person fixed effects included. Regressions are weighted by the probability of selection. Standard errors are clustered at the person level and reported in brackets. Age-in-years, calendar-year, and person fixed effects included in all regressions unless otherwise noted. Standard errors are clustered at the person level and reported in brackets.
$^{***}p < 0.01;^{**}p < 0.05;^{*}p < 0.10$

Appendix Table A.12: Summary statistics for supplementary data sources asking about expectations of future labor supply

	MTF	Gallup	Purdue
White	0.69	0.93	.
Black	0.13	0.07	.
Married	0.10	0.01	.
Expect to be homemaker at 30	0.03	0.60	0.56
Professional w/ PhD or equiv. at 30	0.21	.	.
Professional w/out PhD at 30	0.38	.	.
Def. will get college degree	0.53	0.54	0.21
Choose to have 2 or more kids	0.89	0.97	.
Choose to have 3 or more kids	0.42	0.79	.

Notes: MTF data is from the 12th grade sample of the 1978-2014 Monitoring the Future Surveys. Current marital status and fertility expectations are obtained from Form 3 of the MTF. Education expectations, employment expectations and race are from Form 4 of the MTF. Gallup data is from a 1961 Gallup opinion survey of high school and college students. We retain only respondents in 12th grade. Purdue data is from 1961 and 1962 Purdue Opinion Panels of high school students. Again, we retain only students in 12th grade. Out of worries about anchoring bias, we use only the first wave, in 1961.

Appendix A.2. Data Appendix

A.2.1. BHPS

The BHPS started in 1991 with 5,500 households (10,300 individuals), known as the "original sample." Since then, the sample has grown in various ways. First, the main survey is only administered to adults, so children in original sample households "age in" to the survey upon turning 16. These "agers-in" and all other original sample members are continuously tracked as they form new households, and the individuals in these new households are also interviewed so long as they live with an original sample member. (If these new-comers move out of the household, they are no longer tracked.) One notable exception is that some will "marry in" to the sample and be tracked permanently if they have a child with an original sample member. Children born to original sample members are automatically counted as original sample members.

The BHPS has also grown via the addition of new booster samples. Between 1997 and 2001, the BHPS roughly doubled in size with the addition of 5,000 households (8,600 individuals) out of concerns that the original sample under-represented Scotland, Wales, and Northern Ireland. In 2009, the BHPS became subsumed under the "Understanding Society" dataset, which, including BHPS households, tracked roughly 40,000 households (50,000 individuals) in the UK. While Understanding Society maintained most questions from the BHPS, it notably dropped several of the gender-roles questions we are most interested in, so our main results focus on the BHPS through 2008, for the sake of consistency across our main outcome variables. Results that include data from Understanding Society are not much different from results using BHPS through 2008 only and are available upon request.

Our preferred analysis sample imposes the restriction that we observe each person at least once before becoming a parent and at least once after, so that all observations help identify the main effect of parenthood. Our sample comprises only those who

were original sample members, *i.e.* those who were interviewed in the first wave and *not* those who marry into the sample later on. Because our panel is relatively short (18 waves) compared to our U.S. panel data sources, we exclude children who age into the sample because, in order to meet our restriction that we see the event of the birth of their first child, these individuals would disproportionately have children early (and, related, would be disproportionately female, as women have children at a younger age than men).

The BHPS conducts fertility history interviews for all respondents at the second wave after entry into the sample, and from this set of questions, we can identify parents who have had children before the BHPS starts. Unfortunately, this fertility history is not asked again after the second wave, so we look to the household grid, which lists all individuals who live in a household together as well as some basic demographic information, such as date of birth. Importantly for our purposes, this grid also includes relationships between household members, so we can link children to their parents and construct a full fertility history for each parent in our sample. This is largely the same method suggested by the BHPS. Unfortunately, because a parent must be living with their child in order to identify birth dates, we undercount fathers, as (we hypothesize) that they on average are less likely to be living with their children than are mothers. Relative to a BHPS quality profile conducted at wave 13, we slightly over-count mothers (692 versus 673) and undercount fathers (567 versus 653).

Our main employment outcome variables are defined as follows. Employed last week is equal to 1 if an individual did paid work last week (i.e. was physically present at work) and excludes people who, for example, were on maternity leave. Hours worked is the *typical* hours that an individual works within a given year, and is coded as zero if the individual is not working (including not looking for work). Thus, an individual may be on maternity leave in the week prior to a survey, but their

usual hours worked might be greater than zero, since a respondent can interpret the question to be about when their hours prior to maternity leave. Other employment variables we use include job last week (which includes individuals who had a paid job but for whatever reason were away from it), in labor force currently (which includes individuals who are unemployed and looking for a job), and employed full/part time (which is defined as $\geq$ and < 30 hours, respectively). Finally, we also examine whether an individual's *spouse* is employed in a given year. Note that these individuals are excluded from our sample, as they are not randomly chosen to be included.

We also examine a range of questions on an individuals' opinions on gender roles in the household. Appendix Table A.6 summarizes these questions. While 9 questions are available in total, we only choose a subset of 6 that pertain to the impacts of a woman working. These statements, asked on a scale of 1 (strongly agree) to 5 (strongly disagree) are: (1) A pre school child is likely to suffer if his or her mother works, (2) All in all, family life suffers when the woman has a full time job, (3) A woman and her family would all be happier if she goes out to work, (4) Both the husband and wife should contribute to the household income, (5) Having a full-time job is the best way for a woman to be an independent person, and (6) A husband's jobs is to earn money; a wife's job is to look after the home and family.

We take the individual gender-roles questions and construct an omnibus measure as follows. First, since some questions are phrased in a gender "liberal" manner (i.e. men and women should take on equal roles in the household) and others are phrased in a gender "conservative" manner (i.e. men and women should take on distinct roles in the household), we reverse some questions as necessary so that all are increasing in the gender liberal direction. We then standardize each variable so that they have a mean of zero and standard deviation of one, and sum them up.

Another set of questions we examine in the BHPS relate to expectations of future labor force. The question that we use about labor force expectations is worded as

follows: "Do you think this actually will happen in the coming twelve months? Give up paid work." Note that this question is only asked of individuals who currently have a paid job. We examine this question, combining yes and "don't know," and compare to actual labor force outcomes, conditional on answering the question in the previous wave. These variables are always keyed on the year the expectation was made *about*, so that in event year zero, the expectation variable is made in the year before an individual had their first child. Finally, we construct a variable that asks whether your prediction was correct. This variable (expected to work this year but wrong) is equal to 1 if an individual predicted that she would work this year and was wrong and is equal to 0 if she predicted that she would work this year and was correct. Note that this variable is missing for individuals who expected *not* to work this year. In practice, this group is only 3% of total answers.

A final set of questions we examine in the BHPS relate to job satisfaction. Like the gender-roles questions, we examine each individually and construct an omnibus variable. The individual questions are asked on a scale of 1 (not satisfied at all) to 7 (completely satisfied), and pertain to: (1) overall satisfaction, (2) total pay, (3) security, (4) the work itself, and (5) hours worked. Like the gender-roles questions, we standardize these variables so that they have a mean of zero and standard deviation of one and sum them all together to form an omnibus measure. Note that more questions were available, but they were not asked after wave 7, so for the sake of consistency in our omnibus measure, we selected only those questions that were asked in all waves.

A.2.2. NLSW 1968

Survey Design and Sample

We make use only of the "young women" component of the NLSW68. (While the National Longitudinal Surveys of Young and Mature Women include an older "mature women" sample, they have for the most part had their first child before entering the

sample.) The "young women" component of the NLSW68 is a sample of 5,159 women born between 1941 and 1954. The women were aged 14-24 when first interviewed in 1968 and aged 49-62 when last interviewed in 2003. During the last wave of the survey 55.4% of the original sample was interviewed.

The sample is represented by a multi-stage probability sample drawn by the Census Bureau from 1,900 primary sampling units. In order to provide reliable statistics for black respondents, the NLSW68 oversampled the black population at about twice the expected rate of the total population. Black respondents make up approximately 28 percent of the sample, compared to 11 percent of the population at the time. Probability-of-selection weights created by the NLSW68 correct for this bias.

Sample Selection

Our sample comprises all people in the National Longitudinal Survey of Women who meet the following criteria: (1) We observe them at least once before and after the birth of their first child, (2) We observe them for the first time by age 20 and for the last time after age 40. We include people who drop out at some point in the sample.

In a given year, it is possible to observe a respondent's employment status, but not the number of hours they worked last week, or vise versa. Thus, for a respondent to be included in our sample, we must observe each of the three outcomes variables ("working now", "in labor force now", "hours worked last week").

Employment Status

Employment status is obtained from combining two NLSW68-created variables: employment status recode (ESR), available every 1-3 years from 1968 to 1993, and monthly labor recode (MLR), available every other year from 1995 to 2003. Although not exactly the same, the two variables are very similar. Employment status recode specifies whether the respondent was working, with a job but not at work, laid

off and looking for work, unemployed, going to school, keeping house, unable to work, or retired. Monthly labor recode specifies whether the respondent was employed and at work, employed but absent, unemployed and laid off, unemployed and looking for work, retired, or disabled.

We code respondents as "working now" if they report working (employed and at work) and as "not working" if they report any other activity in the ESR (MLR). Negative values of each variable, which are assigned when respondents refuse to answer, do not know the answer, or are not interviewed, are coded as missing.

We code respondents as "in labor force" if they report working, having a job but not working, looking for work, or being laid off (employed and at work, employed but absent, unemployed and laid off, unemployed and looking for work) and as "not in labor force" if they report any other type of activity in the ESR (MLR). Negative values of each variable, which are assigned when respondents refuse to answer, do not know the answer, or are not interviewed, are coded as missing.

Hours Worked

Between 1968 and 1993, the NLSW68 asks respondents the number of hours they worked in the previous week. We assign this variable to be "hours worked." Respondents deemed to be not working ("working now" = 0) are assumed to have worked zero hours last week.

After 1993, the NLSW68 asks respondents the number of hours worked at their main job and at their other job. We assign our "hours worked" variable to be the sum of these two variables. Again, respondents deemed to be not working ("working now" = 0) are assumed to have worked zero hours last week.

Date of Birth of 1st and 2nd Child

The NLSW68 does not consistently ask respondents for the date of birth of their children. Thus, to determine the date of birth of respondents' first and second children, we combine several categories of variables pertaining to children's date of birth and take the minimum and second minimum value among these.

Between 1968 and 1993, the NLSW68 asks respondents directly about the date of birth of their children. In 1973, the NLSW68 records the date of birth of each child a respondent has had in their lifetime. In 1978, 1983, 1985, 1987, 1988, 1991, and 1993 the NLSW68 records the date of birth of each child the respondent has had since their last interview.

Post 1993, the NLSW68 does not explicitly ask for dates of birth of children, but instead asks for a demographic information, including date of birth, for each member of the household. We include these variables in our calculation of minimum date of birth only if the household member is a child of the respondent (relationship to respondent equal to child, son, or daughter).

To determine the first (second) child's date of birth, we take the minimum (second minimum) date of birth among all date of birth variables described above. Children with missing months of birth but non-missing years of birth are assumed to have been born in September (the most common birth month).

Event Time

The event time variable is calculated by determining the number of months between the date of birth of the respondent's first child and the current interview, dividing by 12, and rounding to the nearest whole number.

Other Demographic Information

Education: Each year, the NLSW68 records the highest completed grade and highest college degree obtained by respondents. We assign the highest grade completed to be the respondent's maximum highest grade reported among all years. Similarly, we assign the highest college degree to be the respondent's maximum reported degree. Respondents who report completing less than 16 years of school are coded as having no college degree.

Age: Age is calculated by determining the number of months between the respondent's date of birth and the current interview, dividing this number by 12, and rounding down. Unknown interview months are assumed to be March.

Mom Worked: In 1968, respondents were asked whether their mother worked when they were 14. Responses of yes are coded as 1, responses of no are coded as 0, and all other values are coded as missing.

Planned to Work at Age 35: In 1968, respondents were asked what they plan to be doing at age 35. Responses of working at a different job and working at the same job are coded as 1. Responses of married, keeping house, other, or do not know are coded as 0. All other values are coded as missing.

A.2.3. NLSY 1979

Survey Design and Sample

The NLSY79 is a sample of 12,686 men and women born between 1957 and 1964. They were aged 14-22 when first interviewed in 1979 and aged 50-57 when last interviewed in 2014.

The NLSY79 purposefully over samples economically disadvantaged, Hispanic, and black youth. Approximately 25 percent of the sample is black and 16 percent of

the sample is Hispanic, compared to their share of population which was around twelve and six percent, respectively, at the time. The inclusion of probability-of-selection weights corrects for this over sampling.

Sample Selection

Our sample comprises all people in the National Longitudinal Survey of Youth who meet the following criteria: (1) We observe them at least once before and after the birth of their first child, (2) We observe them for the first time by age 20 and for the last time after age 40. We include people who drop out at some point in the sample.

In a given year, it is possible to observe a respondent's employment status, but not the number of hours they worked last week, or vise versa. Thus, for a respondent to be included in our sample, we must observe each of the three outcomes variables ("working now", "in labor force now", "hours worked last week"). For regressions on gender-role attitudes, we must observe the respondent's answers to each of the eight questions to be included in the sample.

Employment Status

The NLSY79 records respondents' employment status for each week between 1978 and the date of their last interview. To determine a respondent's current employment status, we use the employment status variable corresponding to week the respondent was interviewed.

We code respondents as "working now" if their employment status is a job number (values of 100 through 3000) as as "not working" if they report not working, being unemployed, being out of the labor force, or being in active military service. All other values are coded as missing.

We code respondents as "in labor force" if their employment status is a job number or they report being unemployed and as "not in labor force" if they report being out of the labor force or in active military service. All other values are coded as missing.

Hours worked

As with employment status, the NLSY79 records the number of hours respondents worked for each week between 1978 and the date of their last interview. We assign this variable, in the week-year corresponding to each interview, to be "hours worked." Respondents deemed to be not working ("working now" = 0) are assumed to have worked zero hours in the previous week.

Gender-role attitudes

NLSY79 asks respondents about their attitudes towards women working four times over the course of the survey – in 1979, 1982, 1987, and 2004. In particular, they ask if respondents personally agree or disagree with each of the following statements

a: A woman's place is in the home, not in the office or shop

b: A wife who carries out her full family responsibilities doesn't have time for outside employment

c: A working wife feels more useful than one who doesn't hold a job

d: The employment of wives leads to more juvenile delinquency

e: It's much better for everyone concerned if the man is the achiever outside the home and the woman takes care of the home and family

f: Men should share the work around the house with women, such as doing the dishes, cleaning, and so forth

g: Women are much happier if they stay at home and take care of their children

We recode each question so that a higher value indicates a more gender-liberal position. Non-interviews as well as responses of "do not know" and "refused" are coded as missing.

Date of Birth of 1st and 2nd Child

The NLSY79 provides the date of birth for each of the respondent's children. We take the minimum (second minimum) date of birth among all children as the birthday of the first (second) child.

Event Time

NLSY79 provides the date of interview for each wave. The event time variable was calculated by determining the number of months between the birth of the respondent's first child and the current interview, dividing by 12, and rounding to the nearest whole number.

Other Demographic Information

Education: Each year, the NLSY79 records the highest completed grade. We assign the highest grade completed to be the respondent's maximum highest grade reported among all years. A respondent is assumed to have a college degree if they have completed 16 or more years of schooling.

Age: Age is calculated by determining the number of days between the respondent's date of birth and the current interview, dividing this number by 365, and rounding down. Unknown interview months are assumed to be March and unknown interview days are assumed to be the 15th.

Mom Worked: In 1979, respondents were asked whether their mother worked when they (the respondents) were 14 years old. Responses of yes are coded as 1, responses of no are coded as 0, and all other values are coded as missing.

Planned to Work at Age 35: In 1979-1984 respondents were asked what they plan to be doing at age 35. Responses of working at a different job and working at the same job are coded as 1. Responses of married, keeping house, other, or do not know are coded as 0. All other values are coded as missing.

A.2.4. PSID

Survey Design and Sample

The PSID started in 1968 with a national sample of about 4,800 U.S. households. This initial sample was made up of two sub-samples. The first was a cross-sectional national sample conducted by the Survey Research Center (SRC) that produced about 2,900 interviews. The second was a sample of low-income families conducted by the Survey of Economic Opportunity (SEO) that produced about 1,900 interviews. The low income sample had unequal selection probabilities as it was limited to SMSAs and non-SMSAs in the southern region. By construction, the PSID oversamples low-income households.

In 1997 and 1999 the PSID added an immigrant booster sample of about 500 households. Those eligible for this sample must have immigrated to the US after 1968 or have been born 1969 or later to people not in the US in 1968.

The PSID includes probability-of-selection weights that are designed to adjust for the oversampling of low-income households and for differential attrition. They are especially important given the combination of SEO and SRC samples that made up the original PSID sample. Weights are also adjusted every five years (1969, 1974, etc) for cumulative panel attrition.

PSID follows sample members no matter their living arrangements: if they split off and form a new household, then that household is added to the sample and each of its members are interviewed; if they have children, then each of their children are interviewed.

Sample Selection

Our sample comprises all people in the Panel Study of Income Dynamics who meet the following criteria: (1) We observe them at least once before and after the birth of their first child, (2) We observe them for the first time by age 20 and for the last time after age 40, and (3) They either begin in or are born into a household in the "core sample" – that is, a household randomly selected to be included in the sample. We include those who are added later on in "booster" samples that increase representation among immigrant and Latino populations and include people who drop out at some point in the sample.

In a given year, it is possible to observe a respondent's employment status, but not the number of hours they worked last week, or vise versa. Thus, for a respondent to be included in our employment sample, we must observe each of the two outcomes variables ("working now" and "in labor force now").

Employment Status

Employment status in the PSID is obtained from the individual employment status variables, available from 1979 to 2015. The variable specifies whether the respondent is working now, temporarily laid off, unemployed and looking for work, retired, permanently disabled, keeping house, or a student.

We code respondents as "working now" if they report working now and coded as "not working" if they report being temporarily laid off, unemployed and looking for

work, retired, permanently disabled, keeping house, or a student. All other values are coded as missing.

We code respondents as "in labor force" if they report working now, being temporarily laid off, unemployed and looking for work. They are coded as "not in labor force" if they report retired, permanently disabled, keeping house, or a student. All other values are coded as missing.

Date of Birth of 1st and 2nd Child

To determine the date of birth of respondent's first and second child, we use the childhood and adoption history file. We limit the data to birth records (as opposed to adoption records), so the date of birth of first and second children correspond only to birth children.

If the child's month of birth is missing, we assume they were born in September. If the child is listed as having been born in winter, spring, summer, or fall we assume they were born in January, April, July, and October, respectively. If the child's birth year is missing, we code the date of birth as missing. The minimum (second minimum) date of birth among all the respondent's births is taken to be the date of birth of the first (second) child.

Event Time

The event time variable is calculated by determining the number of months between the date of birth of the respondent's first child and the current interview, dividing by 12, and rounding to the nearest whole number.

Other Demographic Information

Education: In each survey wave, the PSID records the number of years of education a respondent has received. A respondent is coded as having a college degree if they ever report having completed 16 or more years of schooling.

Age: To determine age, we rely on the PSID variable which asks respondent's for their age at the time of the interview. We then calculated an imputed age – equal to the number of months between the respondent's date of birth and current interview, divided by 12, and rounded down. If the reported age is more than two years from the imputed age or if the current age is lower than a previously reported age, we replace the age variable with our imputed age variable.

Childhood Development Supplement (CDS)

In 1997 PSID began to supplement its main data collection with a survey targeting 0-12 year old children and their parents. Original CDS respondents were interviewed an additional two times, in 2002 and 2007. In 2014, the CDS continued with a new cohort of 0-17 year old children and their parents.

CDS asks children's primary and second caregivers a variety of questions relating to parenthood. Primary caregivers are defined, in order of precedence, as the biological, step, foster, or adoptive mother of the child, the "wife" of a a PSID head who is father to the child, the biological, step, foster, or adoptive father of the child, the legal guardian of the child, and the adult in the household unit who takes primary responsibility for the child. Thus, primary caregivers are almost always women. Secondary caregivers (known as other caregivers after 1997), are defined, in order of precedence, as the biological, step, foster, or adoptive father of the child, the grandmother of the

child, the boyfriend or girlfriend of the primary caregiver, another adult relative of the child, and another adult non-relative of the child.

Our analysis uses the "parenting is hard" question in the CDS. The question asks caregivers to rank on a 5 point scale how much they agree or disagree with the statement "being a parent is harder than I thought it would be." A value of 1 corresponds to "strongly disagree" whereas a value of 5 corresponds to "strongly agree." The "parenting is hard" question is asked to both primary and secondary caregivers in 1997, 2002, and 2007 but only to primary caregivers in 2014.

A.2.5. Supplementary data sources asking expectations of future labor supply

The Monitoring the Future (MTF) survey is an annual cross-sectional survey of American youth. It was first conducted in 1975 and includes about 16,000 high school seniors from approximately 133 public and private high schools in the contiguous United States. Beginning in 1991, the MTF added nationally representative samples of 8th and 10th graders, although we do not make use of them in this paper.

The MTF uses a multi-stage random sampling procedure to create a nationally representative sample of students from each grade. In the first two stages, researchers select particular geographic areas and (with probability proportionate to size) one or more schools in each area. The last stage involves selecting students within each school. In a given school, up to 350 students can be selected. Typically, in schools with fewer than 350 students, all students are included. In larger schools, entire classrooms are randomly sampled.

Our analysis makes use of a question that asks respondents the type of work they expect they will be doing at age thirty. Specifically, the question asks, *"What kind of work do you think you will be doing when you are 30 years old? Mark the one that comes closest to what you expect to be doing."* There are fifteen possible

occupational categories that students can choose from, with "home-maker" as one possible option. Other possible choices include laborer; service worker; operative or semi-skilled worker; sales clerk in a retail store; clerical or office worker; craftsman or skilled worker, sales representative; protective service; manager or administrator; professional without doctoral degree; professional with doctoral degree or equivalent; owner of small business; farm owner/manager; and military service. Below is a list of example occupations provided in the MTF questionnaire for each occupational category.

Job Group	MTF 1976	MTF2016
Laborer	car washer	custodian
	sanitary worker	material mover
	farm laborer	maid
		landscape worker
		farm worker
Service worker		food preparer or food service
	cook	worker including fast food
	waiter	waiter/waitress
	barber	call center worker
	janitor	stock clerk
	gas station attendant	order filler
	practical nurse	nursing aide/orderly
	beautician	teacher assistant
		childcare worker
Operative or semi-skilled worker	garage worker	bus or truck driver
	taxicab	maintenance or repair worker
	bus or truck driver	assembly line worker
	assembly line worker	
	welder	
Sales clerk in a retail store	shoe salesperson	cashier
	department store clerk	supervisor of retail workers
	drugstore clerk	
Clerical or office worker		
	bank teller	secretary
	bookkeeper	receptionist
	secretary	bookkeeper
	typist	supervisor of office workers
	postal clerk or carrier	bank teller
	ticket agent	postal clerk or carrier
Craftsman or skilled worker		
	carpenter	carpenter
	electrician	mechanic
	bric layer	machinist
	mechanic	welder
	machinist	
	tool and die maker	
	telephone installer	

Job Group	MTF 1976	MTF2016
Sales representative	insurance agent real estate broker bond salesman	insurance agent real estate
Protective service	police officer fireman detective	police firefighter paramedic
Manager or administrator	office manager sales manager school administrator government official	office manager government official sales manager
Professional without doctoral degree	registered nurse librarian engineer architect social worker technician accountant actor artist musician	registered nurse school teacher accountant architect artist information technology worker
Professional with doctoral degree or equivalent	lawyer physician dentist scientist college professor	lawyer physician dentist scientist college professor
Owner of small business	restaurant owner shop owner	
Farm owner, farm manager		
Military service		
Full-time homemaker or housewife		

We also make use of the 1961 wave of a two year Purdue Opinion Panel of high school students grades 9-12. We restrict our sample to female high school seniors. The panel asks "What kind of job do you expect to have 20 years from now?" and respondents can choose between twelve possible occupations, including "housewife." Other possible occupations include salesman; owner of factory or small business; professional-teacher, doctor, musician, musician, scientist; office worker; clerk in a store; farm or ranch owner; farm or construction laborer; factory worker or mechanic; big business management; deliveryman, truck driver; and carpenter, plumber, electrician. Raw data and documentation for this survey can be found at `https://ropercenter.cornell.edu/CFIDE/cf/action/catalog/abstract.cfm?type=&start=&id=&archno=USPOP1961-063&abstract=.`

Finally, we use a 1961 Gallup Opinion Poll of young high school, college, and non students. We again restrict our sample to female high school seniors. The Gallup poll asks respondents the open-ended question "What do you expect to be doing when you are 40 years old?" Responses were organized into 18 categories. We use the "house wife," "home maker," "house work," and "raising children" as our definition of expected homemaker. Other categories include teacher, sports

coach, professor; scientist, physicist, biologist, chemist, medical research, psychologist, research; business executive, own business, industry, management, business administration; minister, missionary, social worker; engineering, research engineer, managing engineer; entertainment, actor, broadcasting; medicine, dentist, psychiatry; lawyer; farmer, rancher, agriculture; armed services; government work; sales, clerical: secretary, sales clerk, office worker; nurse; mechanic, machinist, tool and dye maker; skilled trade: electrician, plumber, carpenter, mason, electronics; contracting, building construction, excavating; and artist, cartoonist, designer, draftsman, decorator. Raw data and documentation for this survey can be found at `https://ropercenter.cornell.edu/CFIDE/cf/action/catalog/abstract.cfm?type=&start=&id=&archno=USAIPOSPPOS1961-544&abstract=`.

Appendix A.3. Full model with proofs

A.3.1. Overview

In our model, a young woman forms an estimate of how hard it will be to take care of children while working, which informs her educational decisions. She forms this estimate by observing her own mother, but this estimate is subject to two types of forecasting errors. First, her mother's generation might have higher or lower mean costs of motherhood than her generation. Second, relative to the rest of her cohort, her own mother may have had idiosyncratic (mean-zero) higher or lower costs of motherhood. While the daughter inherits part of her mother's costs, neither of these two error terms gets passed on to her.

Assumptions and set-up

Let utility $u(c, h)$ be a quasi-linear function of consumption c and labor h (for hours worked, say). Specifically, assume that

$$u(c, h) = c - \frac{h^{\gamma+1}}{\gamma + 1},$$

where $\gamma > 0$.

Women's consumption will be equal to market wages net of employment costs of motherhood (both per hour) times labor supply (in hours). Market wages are $\tilde{w} = w + \beta \cdot e$, where w is a base wage for those with higher education $e = 0$ and $\beta > 0$ is the hourly return to higher education $e = 1$. Gaining education $e = 1$ costs some "tuition" $\alpha > 0$, while $e = 0$ is free.

We view "employment costs of motherhood" very broadly, as any cost mothers endure during work. These might include, for example, the per-hour cost of a nanny or day-care service or the emotional or psychic cost of being separated from the child while at work. Actual employment costs for a given woman i are μ_i, but she predicts

this cost with some error δ_i. In particular, she observes her mother's employment costs, $\mu_i + \delta_i$, but she only inherits the μ component. The δ component was her mother's "luck" (good or bad) and is *not* passed down to the daughter. This luck is the sum of a generational component and an idiosyncratic component. For example, her own mother's idiosyncratic good luck might include having had a very understanding supervisor at work, or having parents or in-laws who live in close proximity and can provide free child care. Generational "good luck" could be, for example, giving birth in a moment with cheap, technological substitutes for child care. But the woman in our problem cannot distinguish between μ_i and δ_i and instead uses the sum as the best signal of her own, future employment costs of motherhood.[1]

Assume further that $\mu \sim U[0, 1]$ and $\delta = \{\lambda - \epsilon, \lambda + \epsilon\}$ with equal probability, where λ and $\epsilon > 0$ are both constants. Note that λ can take negative values (for much of the analysis below, λ will drop out, but allowing $\mathbb{E}(\delta) \neq 0$ will be useful later). We assume μ and δ are independent. For simplicity, we further assume no taxes and that w is sufficiently large that $w - \mu - \delta$ is always greater than zero (so mothers' predicted effective wages are always positive).

We abstract away from many possible dimensions of heterogeneity. There is no variation in fertility in our model: women become mothers with probability one and we ignore issues related to delaying or timing childbirth. Unlike many models of human capital, we ignore any variation in "ability" and thus the effective return to and cost of education do not vary along this dimension.

The implicit timing of the model is as follows: A woman makes her education decisions assuming that her future employment costs of motherhood are $\mu + \delta$, but then will make her actual labor supply decisions after this uncertainty is resolved and her actual costs, μ, are revealed to her. While this timing is helpful to keep in

[1]Note also that women in our model are *naive* in the sense that they do not take into account that δ is drawn from a distribution and instead just assume employment costs will be $\mu + \delta$ with probability one.

mind, in fact the problem can be collapsed to a single decision: the woman's (only) problem is to optimally choose education $e \in \{0, 1\}$. Once she chooses e and once her true costs μ are revealed to her, her labor supply is given by a simple optimization problem and thus can be viewed as deterministic. Note that, as we are assuming away many other dimensions of heterogeneity, variation in optimal e comes entirely from variation in expected employment costs, $\mu_i + \delta_i$.

Working backward, the woman calculates her predicted utility for each $e \in \{0, 1\}$. This comparison requires calculating the optimal predicted h^* *given her assumptions about child-care costs* for each e and then seeing which realized utility is higher. For a given e and prediction of child-care cost $\mu_i + \delta_i$, the optimal predicted $\hat{h}$ is given by setting

$$\frac{\partial}{\partial h}\left[(w + \beta e - \mu - \delta)h - \alpha e - \frac{h^{\gamma+1}}{\gamma + 1}\right] = 0,$$

which yields the following *predicted* labor-supply function:

$$\hat{h}^* = (w + \beta e - \mu - \delta)^{\frac{1}{\gamma}}. \tag{1}$$

Note that, as we would expect, predicted hours will increase in wages (and thus education, as education increases wages) and decreases in hourly child-care costs.[2]

The woman takes her optimal predicted $\hat{h}^*$ for each of the two possible levels of education $e \in \{0, 1\}$ and determines which yields higher utility. She thus compares the following two expressions:

$$u(\hat{h}^*(e=0)) = \underbrace{(w - \mu - \delta)}_{\substack{\text{effective wage} \\ \text{if } e=0}} \cdot \overbrace{(w - \mu - \delta)^{\frac{1}{\gamma}}}^{h^*(e=0)} - \frac{\left[(w - \mu - \delta)^{\frac{1}{\gamma}}\right]^{1+\gamma}}{1 + \gamma} \tag{2}$$

[2]As utility is quasi-linear, there is no income effect and thus the substitution effect dominates and labor supply is always a positive function of wages, consistent with a long line of empirical work suggesting that women (relative to men) increase labor supply more in response to increases in effective wages.

and

$$u(\hat{h}^*(e=1)) = \overbrace{(w+\beta-\mu-\delta)}^{\text{effective wage if } e=1} \cdot \overbrace{(w+\beta-\mu-\delta)^{\frac{1}{\gamma}}}^{h^*(e=1)} - \alpha - \frac{\left[(w+\beta-\mu-\delta)^{\frac{1}{\gamma}}\right]^{1+\gamma}}{1+\gamma}. \tag{3}$$

After some algebra, it is easy to show that expression (3) will be greater than (2) (i.e., the woman will choose $e=1$) iff:

$$(w+\beta-\mu-\delta)^{\frac{1+\gamma}{\gamma}} - (w-\mu-\delta)^{\frac{1+\gamma}{\gamma}} > \alpha\frac{1+\gamma}{\gamma}. \tag{4}$$

Intuitively, a higher tuition cost α will discourage choosing $e=1$. As $\frac{1+\gamma}{\gamma} > 0$ (in fact, it is greater than 1) by assumption, the LHS of the equation is increasing in β so more women will choose $e=1$ as the return to education rises (also intuitive).

Less intuitive but important is that the LHS is decreasing in $\mu+\delta$. This claim follows from $\frac{1+\gamma}{\gamma} > 1$ and can be easily shown by taking the derivative of the expression with respect to $\mu+\delta$.

A.3.2. Predictions yielded by the model

Claim 1. *Women who choose $e=1$ will on average work more (post-baby).*

Proof. We first prove the following useful lemma:

Lemma 1. *For a given set of parameter values $(w,\beta,\gamma,\alpha,\lambda,\epsilon)$ for which (the technical condition) $\beta^{\frac{1+\gamma}{\gamma}} < \alpha\frac{1+\gamma}{\gamma}$ and a given realization of the error term $\delta = \{\lambda-\epsilon, \lambda+\epsilon\}$, there exists some μ' such that a woman will choose $e=1$ iff $\mu < \mu'$ when $\delta = \lambda-\epsilon$ and another such cut-off μ'' such that for all $\mu < \mu''$ she will always choose $e=1$ when $\delta = \lambda+\epsilon$.*

Proof. First, fix $\delta = \lambda + \epsilon$. The LHS of equation (4), $(w + \beta - \mu - (\lambda + \epsilon))^{\frac{1+\gamma}{\gamma}} - (w - \mu - (\lambda + \epsilon))^{\frac{1+\gamma}{\gamma}}$, is continuous and strictly decreasing in μ. It also tends to infinity as μ tends to negative infinity. It reaches its minimum (real) value, $\beta^{\frac{1+\gamma}{\gamma}}$, as $\mu \to (w - (\lambda + \epsilon))$. As $\beta^{\frac{1+\gamma}{\gamma}} < \alpha\frac{1+\gamma}{\gamma}$ by assumption, the LHS tends to infinity as $\mu \to -\infty$, then by the intermediate value theorem we know there exists some μ'' such that the LHS and RHS of (4) are equal. Finally, as the LHS is strictly decreasing in μ and the RHS is constant, then this μ'' is unique and equation (4) holds iff $\mu < \mu''$.

A parallel argument establishes a unique value for μ'.

Finally, note that the LHS of equation (4) is everywhere higher when $\delta = \lambda - \epsilon$ than when $\delta = \lambda + \epsilon$, so it will equal $\alpha\frac{1+\gamma}{\gamma}$ for a strictly larger value of μ and thus $\mu'' > \mu'$.

While we have now proved the claim as it is stated, note that we have only shown the existence of some unique $\mu', \mu'' \in \mathbb{R}$, not necessarily $\mu', \mu'' \in [0, 1]$. We will heretofore always assume values for $(w, \beta, \alpha, \gamma, \lambda\epsilon)$ such that $0 < \mu'' < \mu' < 1$ to rule out situations where everyone or no one gets educated. Figure A.10 provides a set of such parameter values and shows a graphical derivation of μ', μ''. □

We now return to the proof of Claim 1. Note that the actual labor supply is made *after* the education decision is made and after true employment costs of motherhood μ are revealed. It follows that, holding e and μ fixed, labor supply is simply:

$$h^* = (w + \beta e - \mu)^{\frac{1}{\gamma}}. \tag{5}$$

The claim requires us to show that $\mathbb{E}(h|e = 1) - \mathbb{E}(h|e = 0) > 0$, which can now be written (using equation 5 and Lemma 1) as:

$$
\begin{aligned}
&\mathbb{E}(h|e=1) - \mathbb{E}(h|e=0) = \\
&\mathbb{E}\Big[(w+\beta-\mu)^{\frac{1}{\gamma}} \mid \Big((\delta = \lambda - \epsilon) \cap (\mu < \mu')\Big) \cup \Big((\delta = \lambda + \epsilon) \cap (\mu < \mu'')\Big)\Big] - \\
&\mathbb{E}\Big[(w-\mu)^{\frac{1}{\gamma}} \mid \Big((\delta = \lambda - \epsilon) \cap (\mu > \mu')\Big) \cup \Big((\delta = \lambda + \epsilon) \cap (\mu > \mu'')\Big)\Big] = \\
&\frac{1}{2} \cdot \mathbb{E}\Big[(w+\beta-\mu)^{\frac{1}{\gamma}} \mid \mu < \mu'\Big] + \frac{1}{2} \cdot \mathbb{E}\Big[(w+\beta-\mu)^{\frac{1}{\gamma}} \mid \mu < \mu''\Big] - \\
&\left(\frac{1}{2} \cdot \mathbb{E}\Big[(w-\mu)^{\frac{1}{\gamma}} \mid \mu > \mu'\Big] + \frac{1}{2} \cdot \mathbb{E}\Big[(w-\mu)^{\frac{1}{\gamma}} \mid \mu > \mu''\Big]\right) > \\
&\left(\frac{1}{2} \cdot \mathbb{E}\Big[(w-\mu)^{\frac{1}{\gamma}} \mid \mu < \mu'\Big] + \frac{1}{2} \cdot \mathbb{E}\Big[(w-\mu)^{\frac{1}{\gamma}} \mid \mu < \mu''\Big]\right) - \\
&\left(\frac{1}{2} \cdot \mathbb{E}\Big[(w-\mu)^{\frac{1}{\gamma}} \mid \mu < \mu'\Big] + \frac{1}{2} \cdot \mathbb{E}\Big[(w-\mu)^{\frac{1}{\gamma}} \mid \mu < \mu''\Big]\right) = \\
&\frac{1}{2}\left(\mathbb{E}\big(w-\mu)^{\frac{1}{\gamma}} \mid \mu < \mu'\big] - \mathbb{E}\big(w-\mu)^{\frac{1}{\gamma}} \mid \mu > \mu'\big]\right) + \\
&\frac{1}{2}\left(\mathbb{E}\big(w-\mu)^{\frac{1}{\gamma}} \mid \mu < \mu''\big] - \mathbb{E}\big(w-\mu)^{\frac{1}{\gamma}} \mid \mu > \mu''\big]\right) > 0.
\end{aligned}
$$

The last step follows because as $\frac{1}{\gamma} > 0$, $(w-\mu)^{\frac{1}{\gamma}}$ is decreasing in μ and thus both terms in the large parentheses are positive. □

Note that to connect this result directly to the empirical work, we need to add slightly more detail to the timing of the model. We have effectively collapsed the model into a single period (women make their education decision based on expected costs of motherhood, and then instantaneously have a child and have their costs revealed to them). If we instead assume that women all have a (perhaps brief) pre-baby period where everyone works some max "full-time" hours (which is roughly what we see in our data), then the claim maps not only into educated women having higher post-baby labor supply, but also exhibiting smaller employment "mommy effects."

The next result states that, even though more educated women have smaller mommy effects with respect to employment, they are nonetheless the most "surprised" by the demands of motherhood.

Claim 2. *Women with* $e = 1$ *underestimate the costs of motherhood more than women with* $e = 0$*. That is,* $\mathbb{E}[\delta \mid e = 1] < \mathbb{E}[\delta \mid e = 0]$.

Proof. : From the Lemma, we know that $e = 1$ iff event a *or* event b occurs, where event a is $(\delta = \lambda + \epsilon) \cap (\mu < \mu'')$ and event b is $(\delta = \lambda - \epsilon) \cap (\mu < \mu')$. Recall that μ and δ are independent.

We can thus write:

$$\mathbb{E}[\delta \mid e = 1] = \mathbb{E}[\delta \mid a]\frac{P(a)}{P(a) + P(b)} + \mathbb{E}[\delta \mid b]\frac{P(b)}{P(a) + P(b)} =$$
$$(\lambda + \epsilon)\frac{P(\delta = \lambda + \epsilon) \cdot P(\mu < \mu'')}{P(\delta = \lambda + \epsilon) \cdot P(\mu < \mu'') + P(\delta = \lambda - \epsilon) \cdot P(\mu < \mu')} +$$
$$(\lambda - \epsilon)\frac{P(\delta = \lambda - \epsilon) \cdot P(\mu < \mu')}{P(\delta = \lambda + \epsilon) \cdot P(\mu < \mu'') + P(\delta = \lambda - \epsilon) \cdot P(\mu < \mu')} =$$
$$(\lambda + \epsilon)\frac{\frac{1}{2}\mu''}{\frac{1}{2}\mu'' + \frac{1}{2}\mu'} + (\lambda - \epsilon)\frac{\frac{1}{2}\mu'}{\frac{1}{2}\mu'' + \frac{1}{2}\mu'} =$$
$$\lambda\left(\frac{\mu'' + \mu'}{\mu'' + \mu'}\right) + \epsilon\left(\frac{\mu'' - \mu'}{\mu'' + \mu'}\right) = \lambda + \epsilon\left(\frac{\mu'' - \mu'}{\mu'' + \mu'}\right).$$

Note that the expression multiplying ϵ is negative, as $\mu'' < \mu'$.

We follow a parallel argument to simplify $\mathbb{E}[\delta \mid e = 0]$. Now, $e = 0$ if either event a' or b' occur, where a' is $(\delta = \lambda + \epsilon) \cap (\mu > \mu'')$ and b' is $(\delta = \lambda - \epsilon) \cap (\mu > \mu')$.

$$\mathbb{E}[\delta \mid e=0] = \mathbb{E}[\delta \mid a']\frac{P(a')}{P(a')+P(b')} + \mathbb{E}[\delta \mid b']\frac{P(b')}{P(a')+P(b')} =$$
$$(\lambda+\epsilon)\frac{P(\delta=\lambda+\epsilon)\cdot P(\mu>\mu'')}{P(\delta=\lambda+\epsilon)\cdot P(\mu>\mu'')+P(\delta=\lambda-\epsilon)\cdot P(\mu>\mu')} +$$
$$(\lambda-\epsilon)\frac{P(\delta=\lambda-\epsilon)\cdot P(\mu>\mu')}{P(\delta=\lambda+\epsilon)\cdot P(\mu>\mu'')+P(\delta=\lambda-\epsilon)\cdot P(\mu>\mu')} =$$
$$(\lambda+\epsilon)\frac{\frac{1}{2}(1-\mu'')}{\frac{1}{2}(1-\mu'')+\frac{1}{2}(1-\mu')} + (\lambda-\epsilon)\frac{\frac{1}{2}(1-\mu')}{\frac{1}{2}(1-\mu'')+\frac{1}{2}(1-\mu')} =$$
$$\lambda+\epsilon\left(\frac{\mu'-\mu''}{2-\mu'-\mu''}\right).$$

Note that the expression multiplying ϵ is positive, as $\mu'' < \mu'$ and $\mu' + \mu'' < 2$ as $\mu', \mu'' \in (0,1)$ by assumption.

It thus follows that $\mathbb{E}[\delta \mid e=1] - \mathbb{E}[\delta \mid e=0] < 0$. As λ drops out, note that the claim holds for any value of λ. □

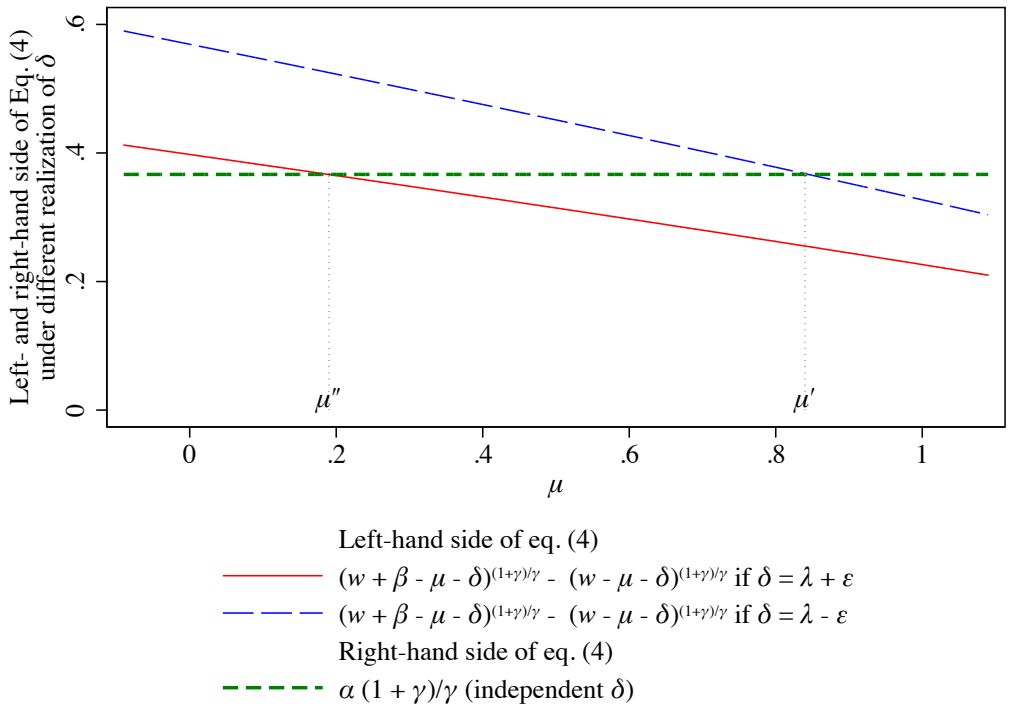
Now, we show that our *unconditional* results on belief-updating (i.e., that, *on average*, women's gender-role attitudes move in the anti-female-employment direction after baby and that they report parenthood being harder than they expected) hold *if and only if* the employment costs of motherhood have increased for the current generation relative to the previous.

Claim 3. $\mathbb{E}[\mu+\delta] < \mathbb{E}[\mu]$ *iff* $\lambda < 0$.

Proof. As $\delta = \{\lambda+\epsilon, \lambda-\epsilon\}$ with equal probability, the claim follows trivially:

$$\mathbb{E}[\mu+\delta] = \mu + \lambda + {}^{1}\!/_{2}\epsilon - {}^{1}\!/_{2}\epsilon = \mu + \lambda < \mu \text{ iff } \lambda < 0.$$ □

Appendix Figure A.10: Graphical derivation of the μ', μ'' cut-off values described in Lemma 1



Notes: To generate this graph, we use the following parameter values:
$w = 2, \gamma = 1.2, \lambda = 0, \epsilon = 0.025, \beta = 0.12, \alpha = 0.2$.

Bibliography

Adda, J., Dustmann, C. and Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, **125** (2), 293–337.

Agresti, A. (2003). *Categorical Data Analysis*. John Wiley & Sons, google-Books-ID: hpEzw4T0sPUC.

Albanesi, S. and Olivetti, C. (2016). Gender roles and medical progress. *Journal of Political Economy*, **124** (3), 650–695.

Alesina, A., Giuliano, P. and Nunn, N. (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, **128** (2), 469–530.

Angelov, N., Johansson, P. and Lindahl, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, **34** (3), 545–579.

Angrist, J. and Evans, W. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, **88** (3), 450–77.

Arum, R., Roksa, J. and Budig, M. (2008). The Romance of College Attendance: Higher Education Stratification and Mate Selection. *Research in Social Stratification and Mobility*, **26**, 107–121.

Attanasio, O., Low, H. and Sánchez-Marcos, V. (2008). Explaining changes in female labor supply in a life-cycle model. *American Economic Review*, **98** (4), 1517–52.

Baxter, J., Buchler, S., Perales, F. and Western, M. (2015). A life-changing event: First births and men's and women's attitudes to mothering and gender divisions of labor. *Social Forces*, **93** (3), 989–1014.

Bertrand, M. (2013). Career, family, and the well-being of college-educated women. *American Economic Review*, **103** (3), 244–50.

— (2018). Coase lecture: the glass ceiling. *Economica*, **85** (338), 205–231.

—, Goldin, C. and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, **2** (3), 228–55.

—, Kamenica, E. and Pan, J. (2015a). Gender identity and relative income within households. *The Quarterly Journal of Economics*, **130** (2), 571–614.

—, — and — (2015b). Gender Identity and Relative Income within Households *. *The Quarterly Journal of Economics*, **130** (2), 571–614.

Blau, F. D. and Kahn, L. M. (2013). Female labor supply: Why is the United States falling behind? *American Economic Review*, **103** (3), 251–256.

Blossfeld, H.-P. and Timm, A. (eds.) (2003). *Who Marries Whom?: Educational Systems as Marriage Markets in Modern Societies*. European Studies of Population, Springer Netherlands.

Blundell, R., Costa Dias, M., Meghir, C. and Shaw, J. (2016). Female labor supply, human capital, and welfare reform. *Econometrica*, **84** (5), 1705–1753.

Borra, C. and Sevilla, A. (2015). *Parental Time Investments in Children: The Role of Competition for University Places in the UK*. Tech. rep., IZA Discussion Paper.

Breen, R. and Salazar, L. (2011). Educational Assortative Mating and Earnings Inequality in the United States. *American Journal of Sociology*, **117** (3), 808–843.

Bronson, M. A. (2014). Degrees are forever: Marriage, educational investment, and lifecycle labor decisions of men and women.

Bruze, G. (2015). Male and female marriage returns to schooling. *International Economic Review*, **56** (1), 207–234.

Bursztyn, L., Fujiwara, T. and Pallais, A. (2017). 'Acting wife': Marriage market incentives and labor market investments. *American Economic Review*, **107** (11), 3288–3319.

Byker, T. (2016). The opt-out continuation: Education, work, and motherhood from 1984 to 2012. *The Russell Sage Foundation Journal of the Social Sciences*, **2** (4), 34–70.

Cancian, M. and Reed, D. (1998). Assessing the Effects of Wives' Earnings on Family Income Inequality. *The Review of Economics and Statistics*, **80** (1), 73–79.

— and — (1999). The Impact of Wives' Earnings on Income Inequality: Issues and Estimates. *Demography*, **36** (2), 173–184.

Cha, Y. and Weeden, K. A. (2014). Overwork and the slow convergence in the gender gap in wages. *American Sociological Review*, **79** (3), 457–484.

Chade, H. and Eeckhout, J. (2017). Stochastic Sorting. *working paper*.

Chen, M. K. and Chevalier, J. A. (2012). Are women overinvesting in education? evidence from the medical profession. *Journal of Human Capital*, **6** (2), 124–149.

Chiappori, P.-A., Salani, B. and Weiss, Y. (2017). Partner Choice, Investment in Children, and the Marital College Premium. *American Economic Review*, **107** (8), 2109–2167.

Chung, Y., Downs, B., Sandler, D. H. and Sienkiewicz, R. (2017). *The Parental Gender Earnings Gap in the United States*. Tech. rep., U.S. Census Bureau Working Paper, CES 17-68.

Currie, J. and Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly journal of economics*, **118** (4), 1495–1532.

DiPrete, T. A. and Buchmann, C. (2006). Gender-specific trends in the value of education and the emerging gender gap in college completion. *Demography*, **43** (1), 1–24.

Doepke, M. and Zilibotti, F. (2017). Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, **85** (5), 1331–1371.

Eika, L., Mogstad, M. and Zafar, B. (2018). Educational Assortative Mating and Household Income Inequality. *Journal of Political Economy*, pp. 000–000.

Fernández, R. (2013). Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review*, **103** (1), 472–500.

Fernández, R., Fogli, A. and Olivetti, C. (2004). Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics*, **119** (4), 1249–1299.

Fernndez, R., Guner, N. and Knowles, J. (2005). Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality. *Quarterly Journal of Economics*, **120** (1), 273–344.

— and Rogerson, R. (2001). Sorting and Long-Run Inequality. *The Quarterly Journal of Economics*, **116** (4), 1305–1341.

Fitzenberger, B., Sommerfeld, K. and Steffes, S. (2013). Causal effects on employment after first birth – a dynamic treatment approach. *Labour Economics*, **25**, 49–62.

Flabbi, L. and Moro, A. (2012). The effect of job flexibility on female labor market outcomes: Estimates from a search and bargaining model. *Journal of Econometrics*, **168** (1), 81–95.

Fortin, N. M. (2005). Gender role attitudes and the labour-market outcomes of women across oecd countries. *Oxford Review of Economic Policy*, **21** (3), 416–438.

Gault, B., Hartmann, H., Hegewisch, A., Milli, J. and Reichlin, L. (2014). *Paid Parental Leave in the United States: What the Data Tell Us about Access, Usage, and Economic and Health Benefits*. Tech. rep., Washington, D.C.: Institute for Women's Policy Research.

Gihleb, R. and Lang, K. (2016). *Educational Homogamy and Assortative Mating Have Not Increased*. Working Paper 22927, National Bureau of Economic Research.

Glassdoor (2016). *Which Countries in Europe Offer the Fairest Paid Leave and Unemployment Benefits?* Tech. rep., Llewellyn Consulting.

Goldin, C. (1990). Understanding the gender gap: An economic history of American women. *Oxford University Press, New York*.

— (2006a). The Quiet Revolution That Transformed Women's Employment, Education, and Family. *AEA Papers and Proceedings*, **May 2006**, 1–21.

— (2006b). The quiet revolution that transformed women's employment, education, and family. *American Economic Review*, **96** (2), 1–21.

— (2014a). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, **104** (4), 1091–1119.

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

—, Katz, L. and Kuziemko, I. (2006a). The Homecoming of American College Women: The Reversal of the Gender Gap in College. *Journal of Economic Perspectives*, **20**, 133–156.

— and Katz, L. F. (2016). *Women Working Longer: Facts and Some Explanations*. Working Paper 22607, National Bureau of Economic Research.

—, — and Kuziemko, I. (2006b). The homecoming of american college women: The reversal of the college gender gap. *Journal of Economic Perspectives*, **20** (4), 133–133.

— and Mitchell, J. (2017). The new life cycle of women's employment: Disappearing humps, sagging middles, expanding tops. *Journal of Economic Perspectives*, **31** (1), 161–82.

Gonalons-Pons, P. and Schwartz, C. R. (2017). Trends in Economic Homogamy: Changes in Assortative Mating or the Division of Labor in Marriage? *Demography*, **54** (3), 985–1005.

Greenwood, J., Guner, N., Kocharkov, G. and Santos, C. (2014). Marry Your Like: Assortative Mating and Income Inequality. *American Economic Review*, **104** (5), 348–353.

—, Seshadri, A. and Yorukoglu, M. (2005). Engines of liberation. *The Review of Economic Studies*, **72** (1), 109–133.

Gregg, P., Gutiérrez-Domènech, M. and Waldfogel, J. (2007). The employment of married mothers in great britain, 1974-2000. *Economica*, **74** (296), 842–864.

Grinza, E., Devicienti, F., Rossi, M. and Vannoni, D. (2017). *How Entry into Parenthood Shapes Gender Role Attitudes: New Evidence from Longitudinal UK Data*. Tech. rep., Working Paper.

Guryan, J., Hurst, E. and Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic Perspectives*, **22** (3), 23–46.

Hellevik, O. (2007). 'Margin Insensitivity' and the Analysis of Educational Inequality. *Sociologick asopis / Czech Sociological Review*, **43** (6), 1095–1119.

Herr, J. L. (2015). The labor supply effects of delayed first birth. *American Economic Review*, **105** (5), 630–37.

— (forthcoming). Measuring the effect of first birth timing on women's wages. *Journal of Population Economics*.

Hotz, J. V., Johansson, P. and Karimi, A. (2017). *Parenthood, family friendly firms, and the gender gaps in early work careers*. Tech. rep., NBER Working Paper No. 24173.

Hou, F. and Myles, J. (2008). The Changing Role of Education in the Marriage Market:Assortative Marriage in Canada and the United States since the 1970s. *Canadian Journal of Sociology*, **33** (2).

Jarrallah, Y., Perales, F. and Baxter, J. (2016). *Parenthood and men's and women's gender-role attitudes: Does child's gender matter?* Tech. rep., Working Paper.

Kalmijn, M. (1991). Shifting Boundaries: Trends in Religious and Educational Homogamy. *American Sociological Review*, **56** (6), 786–800.

Karoly, L. A. and Burtless, G. (1995). Demographic change, rising earnings inequality, and the distribution of personal well-being, 19591989. *Demography*, **32** (3), 379–405.

Kaufmann, K. M., Messner, M. and Solis, A. (2015). *Elite higher education, the marriage market and the intergenerational transmission of human capital*. Tech. rep., Working paper.

Kleven, H., Landais, C. and Søgaard, J. (2016). The breadwinner notch: A bunching approach to estimating reference-dependent preferences. *Unpublished manuscript, London School Econ*.

Kleven, H. J., Landais, C. and Søgaard, J. E. (2018). *Children and gender inequality: Evidence from Denmark*. Tech. rep., NBER Working Paper No. 24219.

Kremer, M. (1997). How Much Does Sorting Increase Inequality? *Quarterly Journal of Economics*, **112** (1), 115–139.

Lee, J. Y. (2018). The upward trend in womens college-going: the role of teenagers anticipated future labour force attachment. *Applied Economics*, **50** (30), 3271–3284.

Lefgren, L. and McIntyre, F. (2006). The relationship between women's education and marriage outcomes. *Journal of Labor Economics*, **24** (4), 787–830.

Liu, H. and Lu, J. (2006). Measuring the degree of assortative mating. *Economics Letters*, **92** (3), 317–322.

Logan, J. A. (1996). Rules of Access and Shifts in Demand: A Comparison of Log-Linear and Two-Sided Logit Models. *Social Science Research*, **25** (2), 174–199.

Logan, T. D. and Parman, J. M. (2017). The National Rise in Residential Segregation. *The Journal of Economic History*, **77** (1), 127–170.

Lordan, G. and Pischke, J.-S. (2016). *Does Rosie like riveting? Male and female occupational choices*. Tech. rep., National Bureau of Economic Research.

Lundborg, P., Plug, E. and Rasmussen, A. W. (2017). Can women have children and a career? iv evidence from ivf treatments. *American Economic Review*, **107** (6), 1611–37.

Mare, R. D. (1991). Five Decades of Educational Assortative Mating. *American Sociological Review*, **56** (1), 15–32.

— (2016). Educational Homogamy in Two Gilded Ages: Evidence from Inter-generational Social Mobility Data. *The ANNALS of the American Academy of Political and Social Science*, **663** (1), 117–139.

Mas, A. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, **107** (12), 3722–59.

McGinn, K. L., Mayra, R. C. and Lingo, E. L. (2015). *Mums the word! Cross-national Effects of Maternal Employment on Gender Inequalities at Work and at Home*. Tech. rep., Harvard Business School Working Paper No. 15-094.

Miller, A. R. (2011). The effects of motherhood timing on career path. *Journal of Population Economics*, **24** (3), 1071–1100.

Novosad, P. and Rafkin, C. (2019). Mortality Change Among Less Educated Americans. *Working paper*.

Olivetti, C., Patacchini, E. and Zenou, Y. (2016). *Mothers, peers and gender identity*. Tech. rep., Working Paper.

Oreopoulos, P. and Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives*, **25** (1), 159–84.

Pertold-Gebicka, B., Petold, F. and Datta Gupta, N. (2016). *Employment adjustments around childbirth*. Tech. rep., IZA Discussion Paper No. 9685.

Qian, Z. and Preston, S. H. (1993). Changes in American Marriage, 1972 to 1987: Availability and Forces of Attraction by Age and Education. *American Sociological Review*, **58** (4), 482–495.

Ramey, G. and Ramey, V. A. (2010). The rug rat race. *Brookings Papers on Economic Activity*.

Reardon, S. F. and Bischoff, K. (2011). Income inequality and income segregation. *AJS; American journal of sociology*, **116** (4), 1092–1153.

Rossin-Slater, M. (2017). Signing Up New Fathers: Do Paternity Establishment Initiatives Increase Marriage, Parental Investment, and Child Well-Being? *American Economic Journal: Applied Economics*, **9** (2), 93–130.

Schank, H. and Wallace, E. (2018). *The Ambition Decisions: What Women Know About Work, Family, and the Path to Building a Life*. Viking.

Schwartz, C. R. (2010). Earnings Inequality and the Changing Association between Spouses Earnings. *AJS; American journal of sociology*, **115** (5), 1524–1557.

— (2013). Trends and Variation in Assortative Mating: Causes and Consequences. *Annual Review of Sociology*, **39** (1), 451–470.

— and Mare, R. D. (2005). Trends in Educational Assortative Marriage from 1940 to 2003. *Demography*, **42** (4), 621–646.

— and — (2012). The Proximate Determinants of Educational Homogamy: The Effects of First Marriage, Marital Dissolution, Remarriage, and Educational Upgrading. *Demography*, **49** (2), 629–650.

Shafer, E. F. and Malhotra, N. (2011). The effect of a child's sex on support for traditional gender roles. *Social Forces*, **90** (1), 209–222.

Shafer, K. and Qian, Z. (2010). Marriage Timing and Educational Assortative Mating. *Journal of Comparative Family Studies*.

Van Bavel, J., Schwartz, C. R. and Esteve, A. (2018). The Reversal of the Gender Gap in Education and Its Consequences for Family Life. *Annual Review of Sociology*, **44** (1), 341–360.

Vespa, J. (2009). Gender ideology construction: A life course and intersectional approach. *Gender and Society*, **23** (3), 363–387.

Waldfogel, J. (1997). The effect of children on women's wages. *American Sociological Review*, pp. 209–217.

Wasserman, M. (2018). *Hours constraints, occupational choice, and gender: Evidence from medical residents*. Tech. rep., Working Paper.

Wilde, E. T., Batchelder, L. and Ellwood, D. T. (2010). *The mommy track divides: The impact of childbearing on wages of women of differing skill levels*. Tech. rep., NBER Working Paper No. 16582.

Wiswall, M. and Zafar, B. (2017). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, **133** (1), 457–507.

Xie, Y. and Killewald, A. (2013). Intergenerational Occupational Mobility in Great Britain and the United States since 1850: Comment. *American Economic Review*, **103** (5), 2003–2020.