(Non-)Marital Assortative Mating and the Gender Gap in Education*

Jenny Shen†

March 31, 2020

Abstract

Educational assortative mating has important implications for household income inequality and intergenerational mobility. Previous work has found that assortative mating has increased over time, using a measurement paradigm that compares observed matches to a counterfactual world where men and women match randomly. I show that the measured increase in assortative mating over the last 50 years is in fact mechanically driven by the closing of the gender gap in education. That is, as men's and women's education distributions grew more similar over time, men and women had more opportunity to find a match with someone of the same education level. Given this finding, I propose a new measure of assortative mating, which I call the Perfect-Random normalization, which bounds observed matches from below by a counterfactual world where men and women match randomly and from above by a counterfactual world where men and women match perfectly according to education. Once I control for the changing gender gap in education using my proposed Perfect-Random normalization, I find that assortative mating has actually *decreased* over time, until around 2000, at which point it began to sharply increase. I then use the Perfect-Random normalization to measure trends in assortative mating among *all* new parents using an administrative dataset of birth certificates in the U.S. Existing work has primarily studied married or cohabiting couples due to data constraints, but births to unmarried parents comprise roughly 40 percent of all births and are concentrated among parents with lower levels of education. I find that sorting among all new parents has been higher than sorting among married parents only since 1990.

*I am deeply grateful to my advisors, Ilyana Kuziemko and Janet Currie, for their encouragement, guidance, and support throughout this project. I also benefitted greatly from conversations and advice from Jesse Bruhn, Emily Cuddy, Kathy Edin, Thomas Fujiwara, Richard Rogerson, Ebonya Washington, Owen Zidar, and Esmée Zwiers, as well as participants in Princeton's Industrial Relations, Fellowship of Woodrow Wilson Scholars, and Center for Health and Wellbeing student lunch seminars.

†Department of Economics, Princeton University, Email: jenny.shen@princeton.edu

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; Fernández and Rogerson, 2001; Fernández *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 mating. In this paper, I develop new tools to do so. Measuring educational assortative mating 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 mating, 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 mating 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 mating 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), 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 mating 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 mating, 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

[1]Because observed matches have been more assortative than the random matching counterfactual, I focus on measuring positive assortative mating. An analogous measure can be defined for negative assortative mating, 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.

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.

Using the Perfect-Random normalization and standard survey data sources (the Current Population Survey), I find that educational assortative mating 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 mating 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 mating 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 mating 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 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 2b, I provide suggestive evidence that assortative mating 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 mating literature. Existing work has used various measurement techniques to find that educational assortative mating 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 mating 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 mating 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 mating has increased over the past 50 years by exploring factors behind this rise. Some have interpreted the measured increase in assortative mating 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 mating 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 mating, 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.

2 Background: Measuring assortative mating

2.1 Measurement Challenges

One reason researchers are interested in studying assortative mating 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 mating 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 certain types of people

[2]This is a particularly lively debate in work estimating 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).

together). We can conceptualize assortative mating in many different ways—for example, a natural starting point for measuring educational assortative mating 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 mating?", researchers tend to implicitly refer to changes in assortative mating 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.*, 2006; Goldin, 2006, 2014; 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 mating should be able to directly measure latter while controlling for the former. What researchers have relied upon 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

[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 mating (Chiappori *et al.*, 2017; Chade and Eeckhout, 2017).

take varying forms but all share the common feature of comparing to the same reference distribution of matches.

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.

2.2.1 Eika *et al.* (2018)

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

$$\text{Random matching normalization} = \frac{Pr(\text{Observed match})}{Pr(\text{Seeing same match under random matching})} \quad (1)$$

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:

$$\text{Random matching normalization} = \frac{Pr(E_m = e \text{ and } E_w = e')}{Pr(E_m = e) \cdot Pr(E_w = e')} \tag{2}$$

Where the denominator comes from assuming that E_m and E_w are statistically independent. They calculate aggregate trends in assortative mating 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 mating and values less than 1 indicate negative assortative mating.

2.2.2 Schwartz and Mare (2005)

Schwartz and Mare (2005), as well as several other papers[4], measure educational assortative mating 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, statistical independence means that:

$$\pi_{ij} = A \cdot \alpha_i \cdot \beta_j \tag{3}$$

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

$$\log(\pi_{ij}) = A' + \alpha_i' + \beta_j' \tag{4}$$

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 4 captures how assortative

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

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)).

2.2.3 Greenwood *et al.* (2014)

Greenwood *et al.* (2014) compare observed matches to random matching in their analysis of whether assortative mating 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.

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

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

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

only changes in the educational similarity of spouses are *mechanical*, and, thus, the random matching normalization should not measure any changes in educational assortative mating 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 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 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

[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.

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 4 together with the measured trends in assortative mating found in Eika *et al.* (2018). Because matching by construction never changes in my simulation exercise, an ideal statistic to measure assortative mating 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 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 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 5b). In fact, the measured trends in my simulation exercise closely resemble measured trends in the "real world" (Figure 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

[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).

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

4 The Perfect-Random Normalization

Given the results of my simulation exercise, I propose a new measure of assortative mating, 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 concretly, my normalization is:

$$\text{Perfect-Random Normalization} = \frac{\text{Observed} - \text{Random}}{\text{Perfect} - \text{Random}} \tag{5}$$

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 6, I plot the share of high school dropout-high school dropout couples as predicted under the

[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.

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 6a) is the distance between the observed and random matching lines, and the denominator of the perfect-random normalization (Figure 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:

$$s(e, e') = \frac{p_{\text{observed}}(e, e') - p_{\text{random}}(e, e')}{p_{\text{perfect}}(e, e') - p_{\text{random}}(e, e')} \tag{6}$$

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 mating, 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).

[10] During the period observed in my data (1962-present), assortative mating 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 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 mating.

[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.

5 Results

5.1 Main results

Figure 7 plots aggregate trends in assortative mating as measured by the Perfect-Random normalization. In contrast to existing work that has found that assortative mating has increased over time, I find that educational assortative mating 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 mating after 2000 driven by the gender gap re-emerging as women have surpassed men in educational attainment.

Looking at component parts in Figure 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).

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 sim $\in$ {Observed, Perfect, 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:

$$\text{Perfect-Random}_{\text{corr}} = \frac{\text{corr}_{\text{observed}} - \cancelto{cov(E_m,E_w)=0}{\text{corr}_{\text{random}}}}{\text{corr}_{\text{perfect}} - \cancelto{cov(E_m,E_w)=0}{\text{corr}_{\text{random}}}} = \frac{\text{corr}_{\text{observed}}}{\text{corr}_{\text{perfect}}} \tag{7}$$

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:

$$\text{Perfect-Random}_{\text{share same}} = \frac{\mathbb{E}[\mathbb{1}[\text{Same}]_{\text{observed}}] - \mathbb{E}[\mathbb{1}[\text{Same}]_{\text{random}}]}{\mathbb{E}[\mathbb{1}[\text{Same}]_{\text{perfect}}] - \mathbb{E}[\mathbb{1}[\text{Same}]_{\text{random}}]} \tag{8}$$

Figure 9 plots trends in educational assortative mating 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 mating 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.

5.3 Different ways of measuring education

One critique of existing work measuring educational assortative mating is that trends in assortative mating are sensitive to how education is measured (Gihleb and Lang, 2016).

In Figure 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 mating, I rely on the weighted average of the Perfect-Random normalization calculated for each combination of husband's/wife's education (equation (4)). Figure 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 mating, 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 mating through 2000 than combining those two categories (i.e., by measuring education using 5 vs. 4 categories). Assortative mating as measured using 2 categories (i.e., "skilled" vs. "unskilled") leads to a finding that assortative mating overall is much higher than as measured using more granular categorizations.

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 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 match with each other. Figure 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.

[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.

6 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 mating 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 mating 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.[14] 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 reported parental education.[15] 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 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

[14]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 Appendix Figure A.2 for data availability and reporting of parental education.

[15]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.

the max years and max states samples, although the max years sample is slightly whiter.

6.1 Main comparison: CPS vs. births

Figure 12 plots trends in educational assortative mating 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 does result in a steeper rise in assortative mating since 2000 (Appendix Figure A.3). Now, the rise in assortative mating after 2000 is steeper than the decline through 2000.

6.2 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 (Appendix Figure A.4; Rossin-Slater (2017)). These mothers are particularly disadvantaged (Table 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 13. When I categorize missing dads as less than high school (rather than dropping these parents from the sample), I find higher levels of assortative mating 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.

6.3 Trends by race

In Figure A.5, I plot trends in assortative mating 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 A.6). In Figure A.7, 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.

6.4 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 A.8).

7 Conclusion

In this paper, I developed new tools to measure educational assortative mating 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 mating, 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 mating 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 mating 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 mating 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 mating 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 mating using information on parental education in the Vital Statistics Birth Certificates dataset. Existing work measuring assortative mating has focused on married or cohabiting couples because of 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 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 mating because of its potential implications for intergenerational mobility. I found that trends in assortative mating 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 mating, 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 mating 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.*, 2015). 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 mating 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 intergenerational occupational or educational mobility are similarly driven by a changing gap in occupation or education distributions between father and son generations.

References

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

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.

Bertrand, M., Kamenica, E. and Pan, J. (2015). Gender Identity and Relative Income within Households *. *The Quarterly Journal of Economics*, **130** (2), 571–614.

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.

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

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.

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

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

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

Fernández, 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.

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

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

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

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

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.

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

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).

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, 1959–1989. *Demography*, **32** (3), 379–405.

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

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.

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.

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

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.

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.

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.

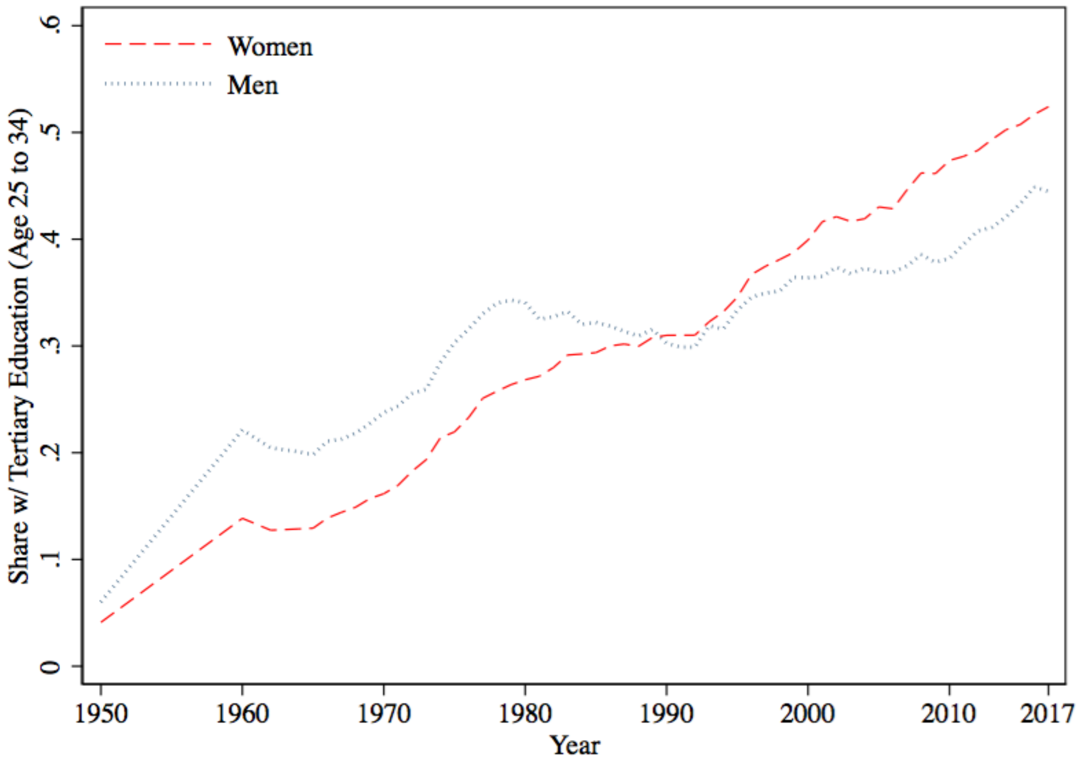
Shafer, K. and Qian, Z. (2010). Marriage Timing and Educational Assortative Mating. *Journal of Comparative Family Studies*.

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.

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.

Figures

Figure 1: Educational attainment of men and women, 1950-current

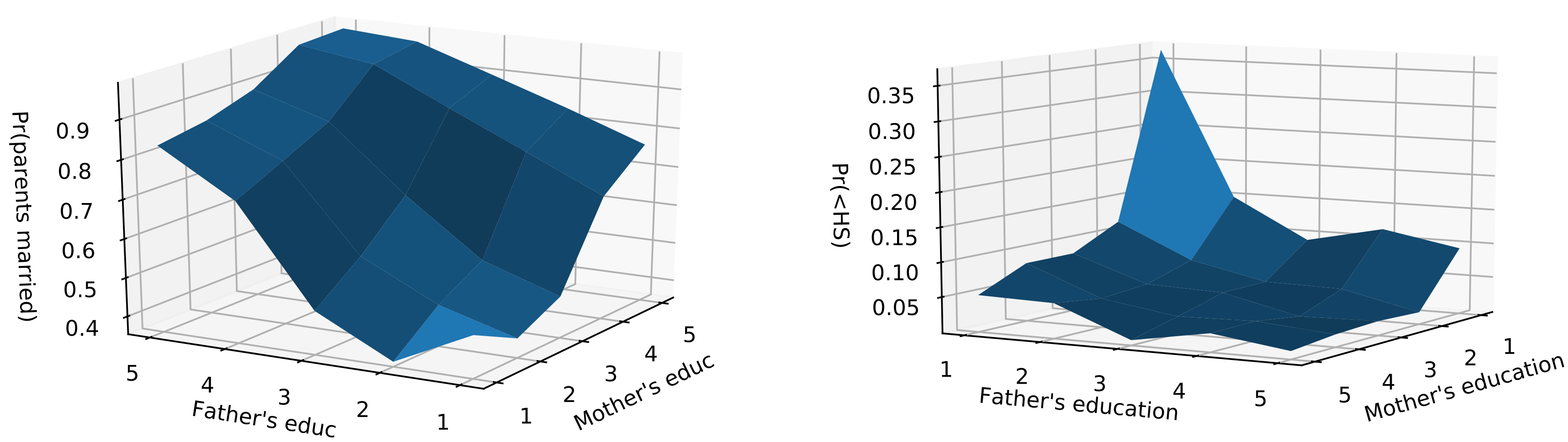


Source: Decennial Census (1950, 1960), Current Population Survey (1962-2017)
Notes: Sample is all men and women aged 25 to 34, and tertiary education is defined as an Associate's degree or above.

Figure 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 3: Perfect Matching Simulation Exercise, Visual Representation

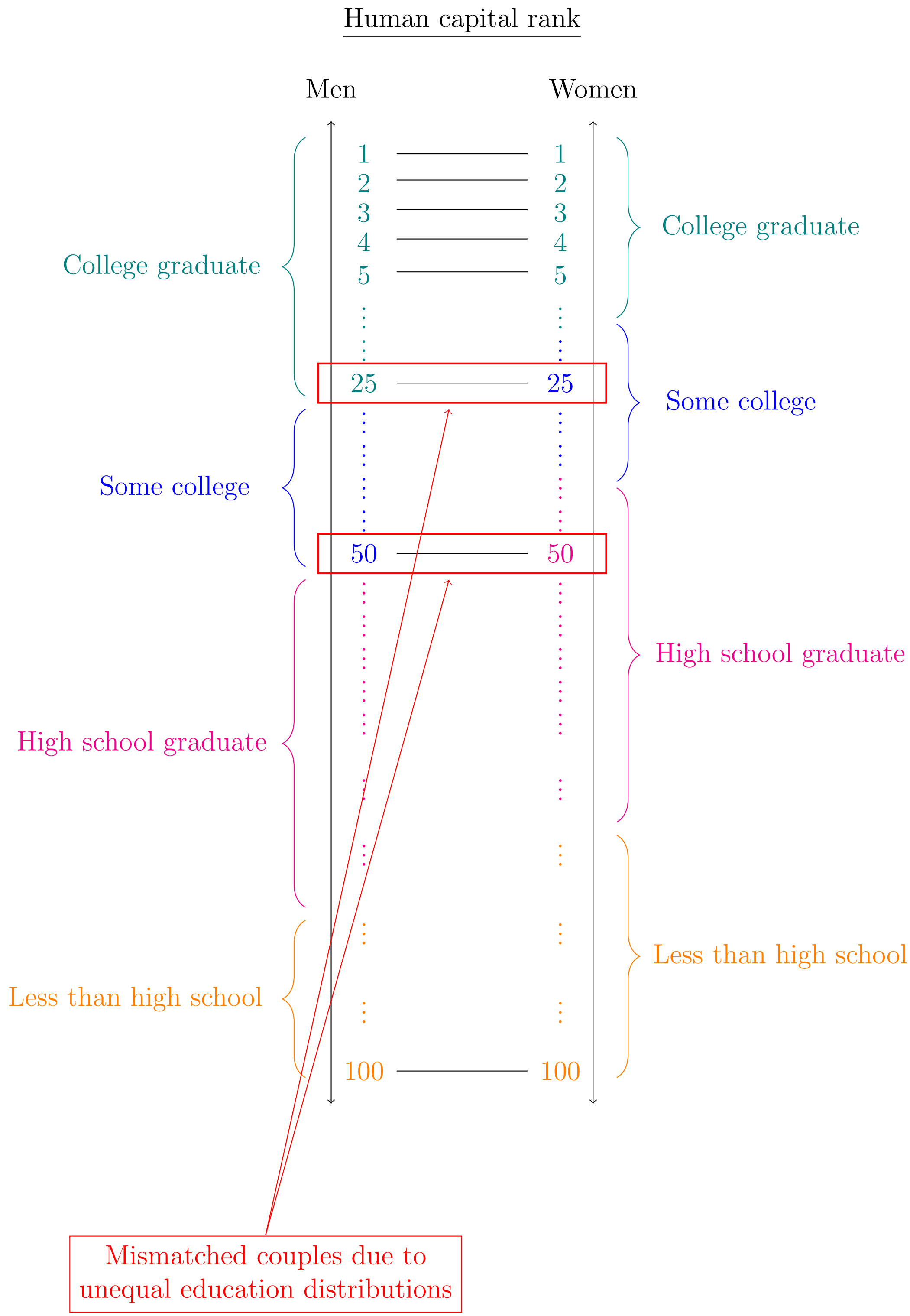
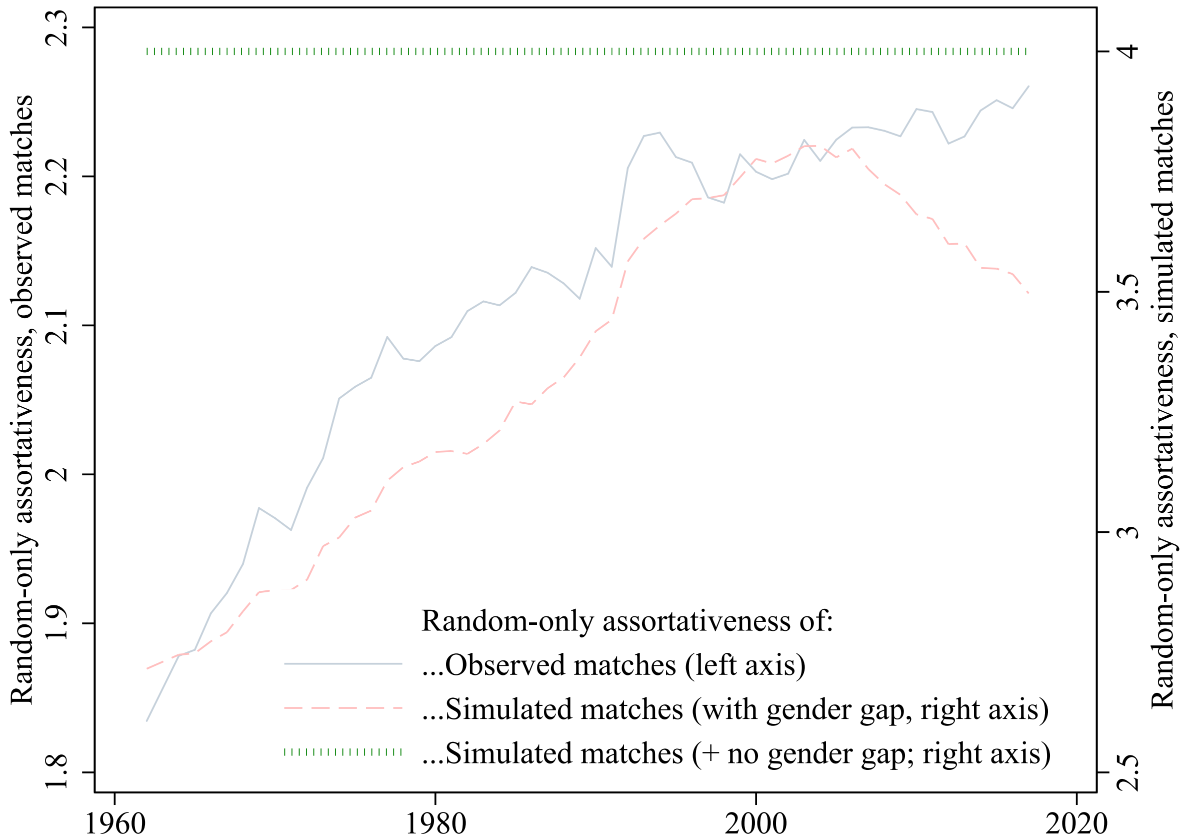


Figure 4: Trends in assortative mating according to the random-only normalization, observed matches vs. simulated (perfect) matches

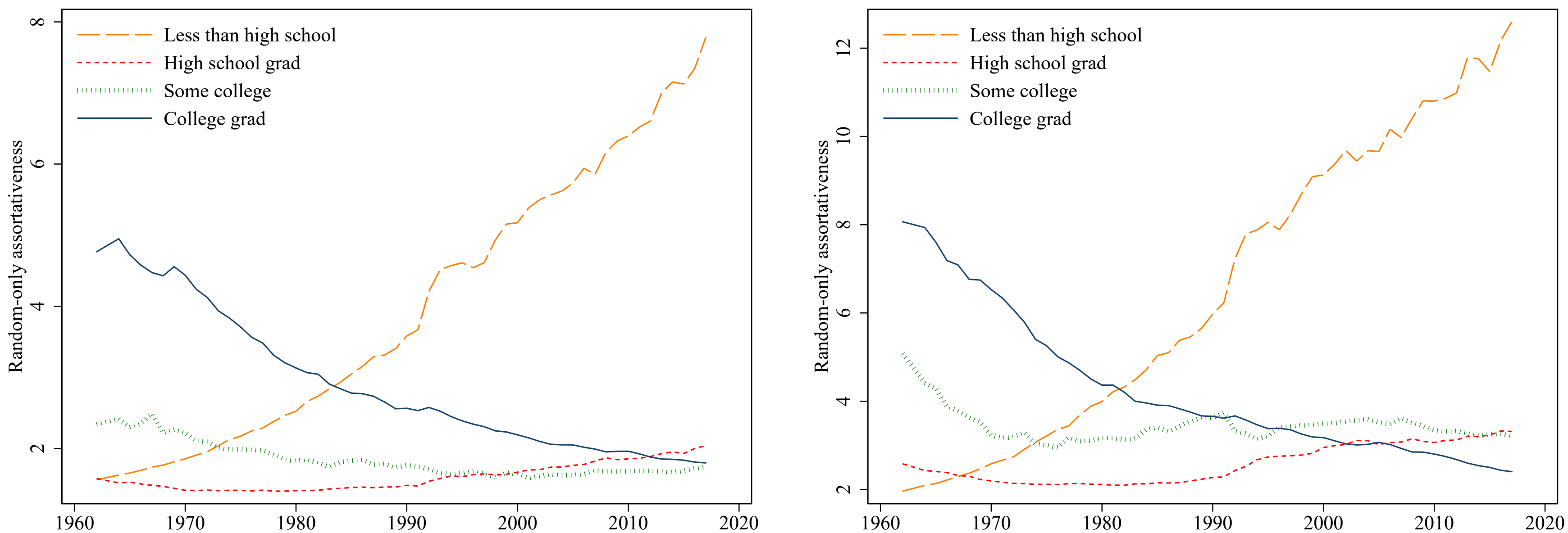


Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some 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 5: Trends in educational assortative mating according to the random-only normalization, split by education category

(a) Observed matches

(b) Simulated (perfect) matches



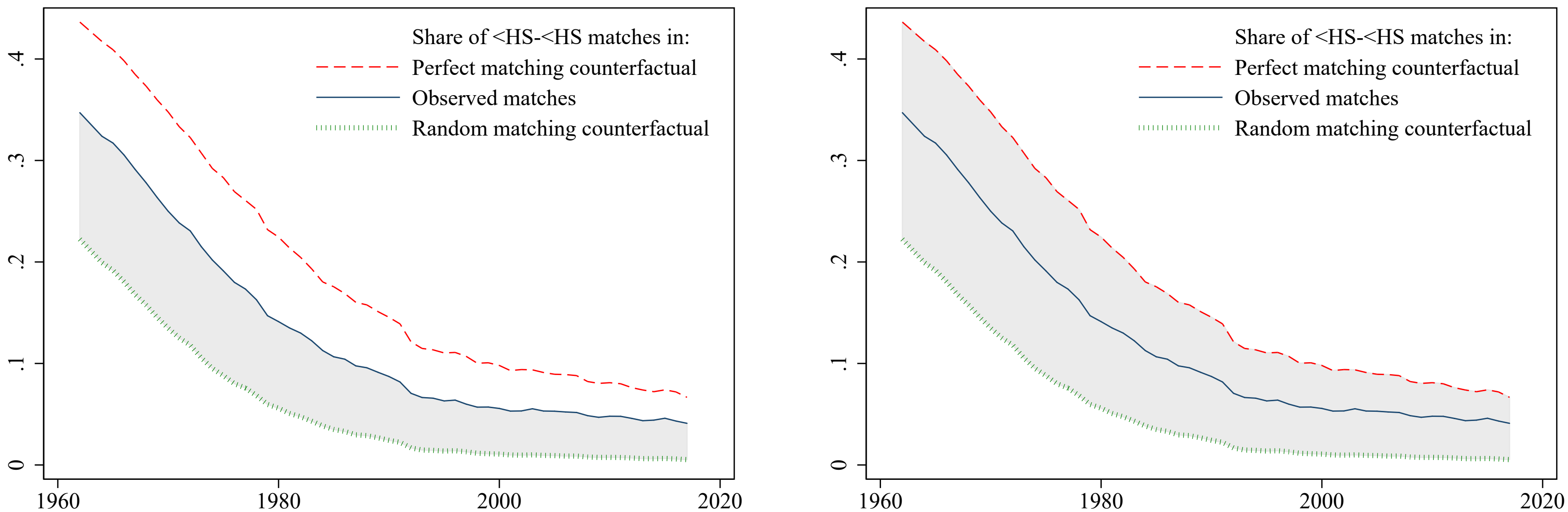
Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some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

Figure 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

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

(a) Numerator = Observed − Random

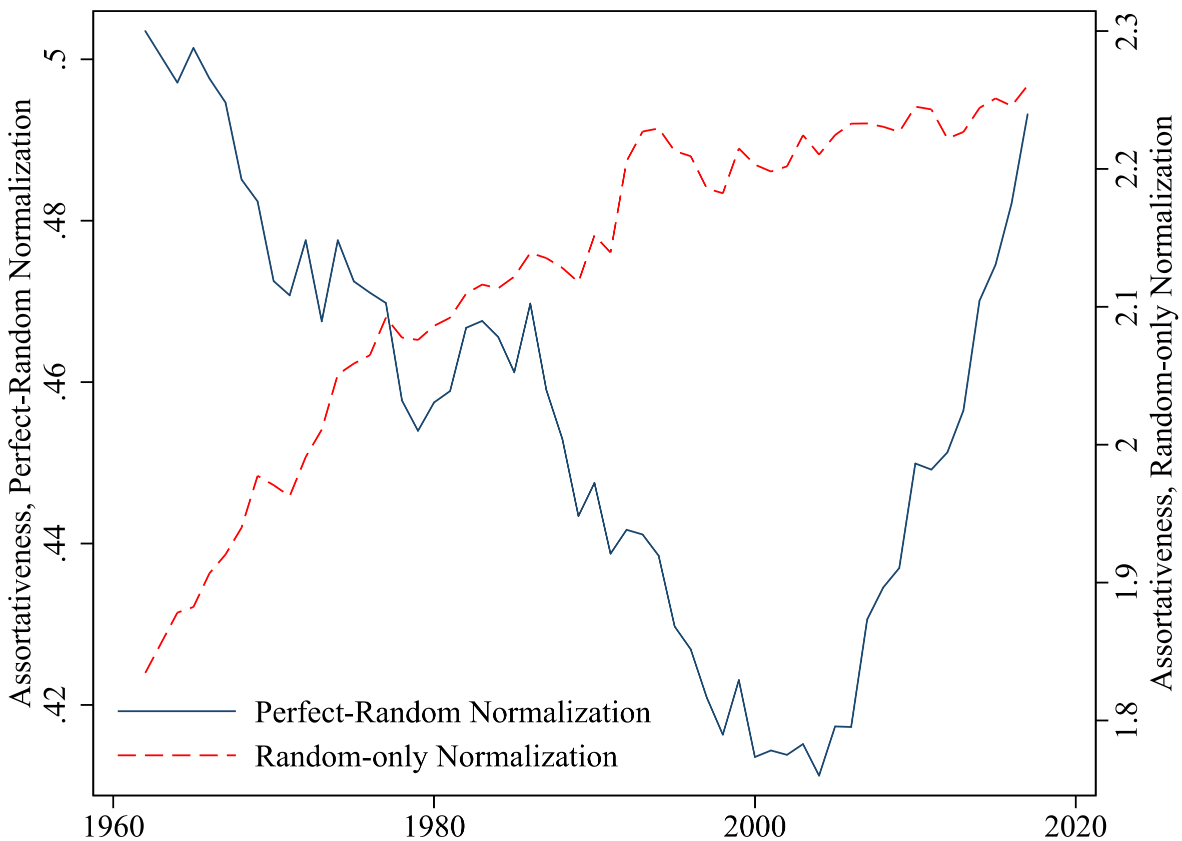
(b) Denominator = Perfect − Random



Source: Current Population Survey

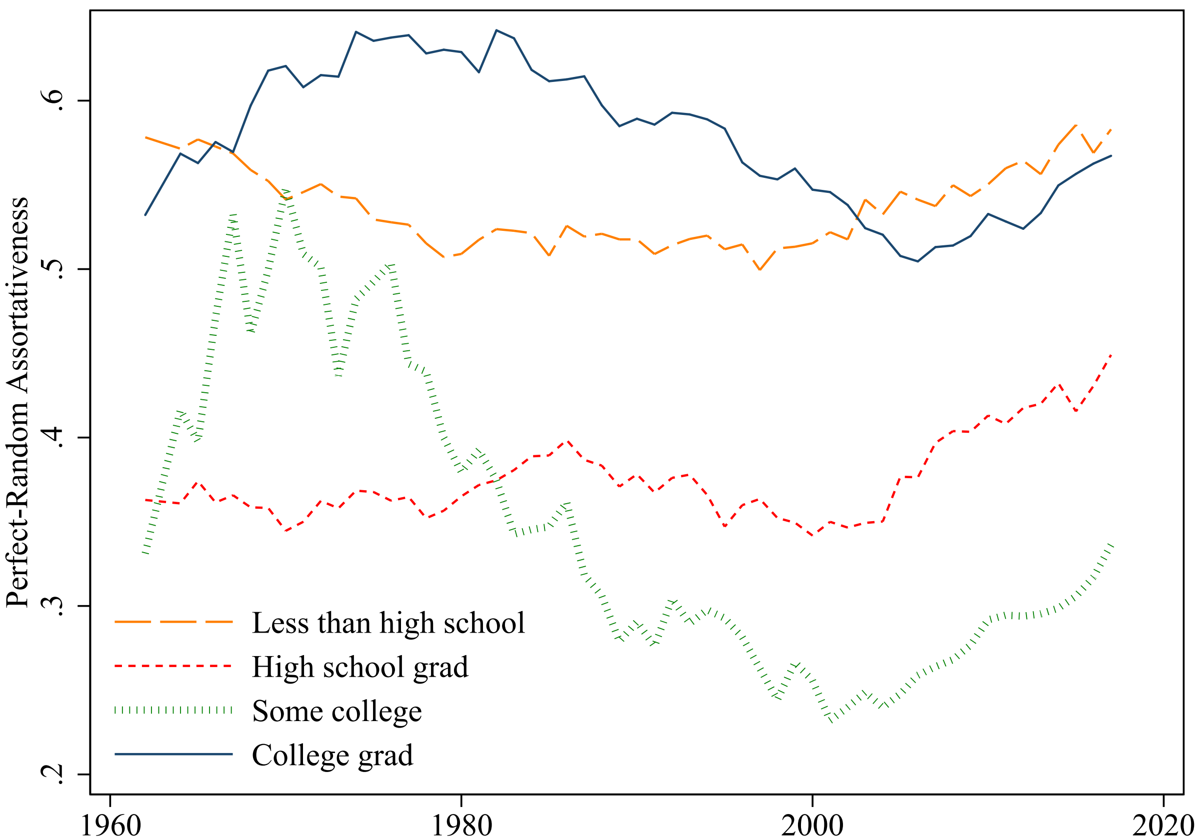
Notes: Sample is all married couples where at least one spouse is between the ages of 26-60, following Eika et al. (2018). Each line plots the share of couples where both partners have 12 years of education / is a high school graduate. The random matching counterfactual assumes that men and women match 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 7: Aggregate trends in educational assortative mating 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 8: Trends in educational assortative mating according to the Perfect-Random Normalization, by category

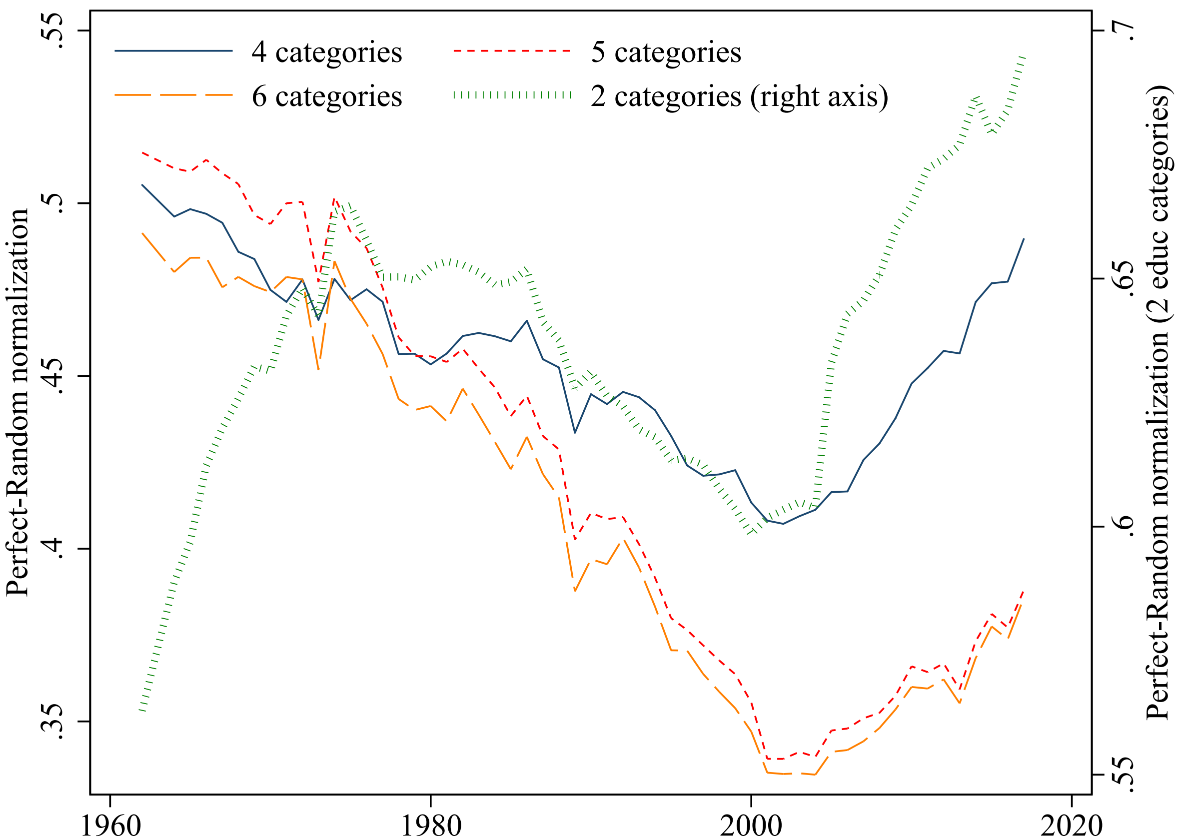


Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some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

Figure 9: Perfect-Random Normalization, Other Statistics

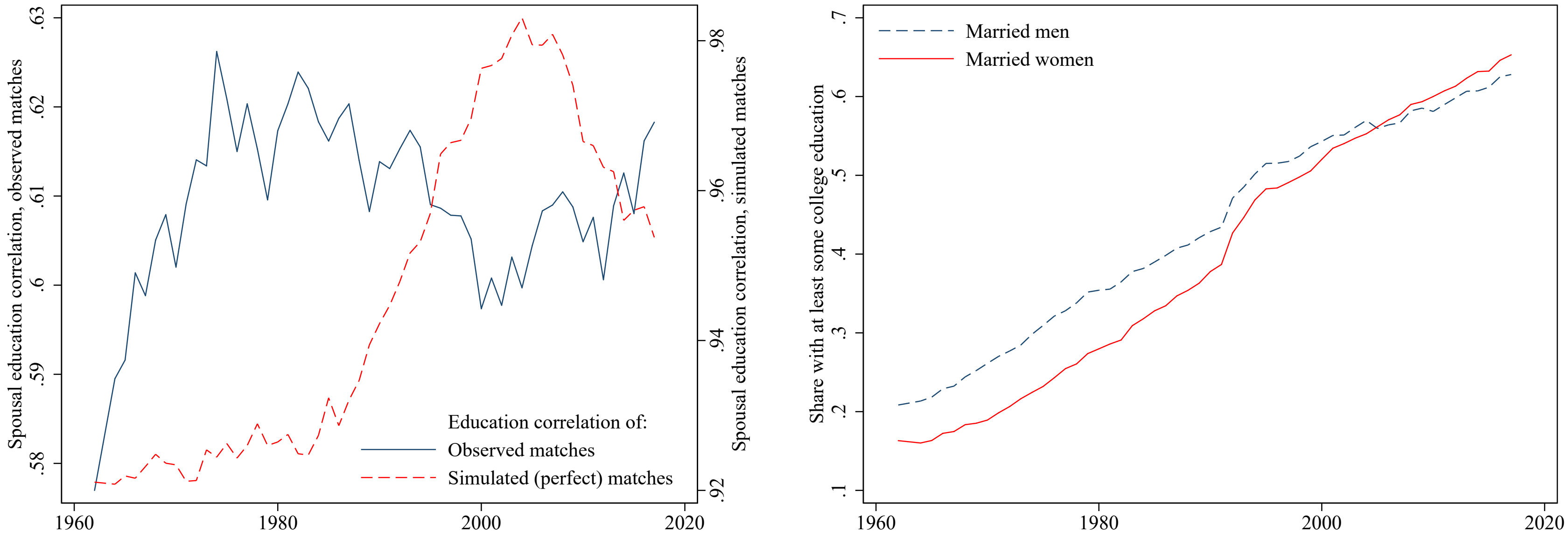
Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some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

Figure 10: Perfect-Random Normalization, various education categorizations



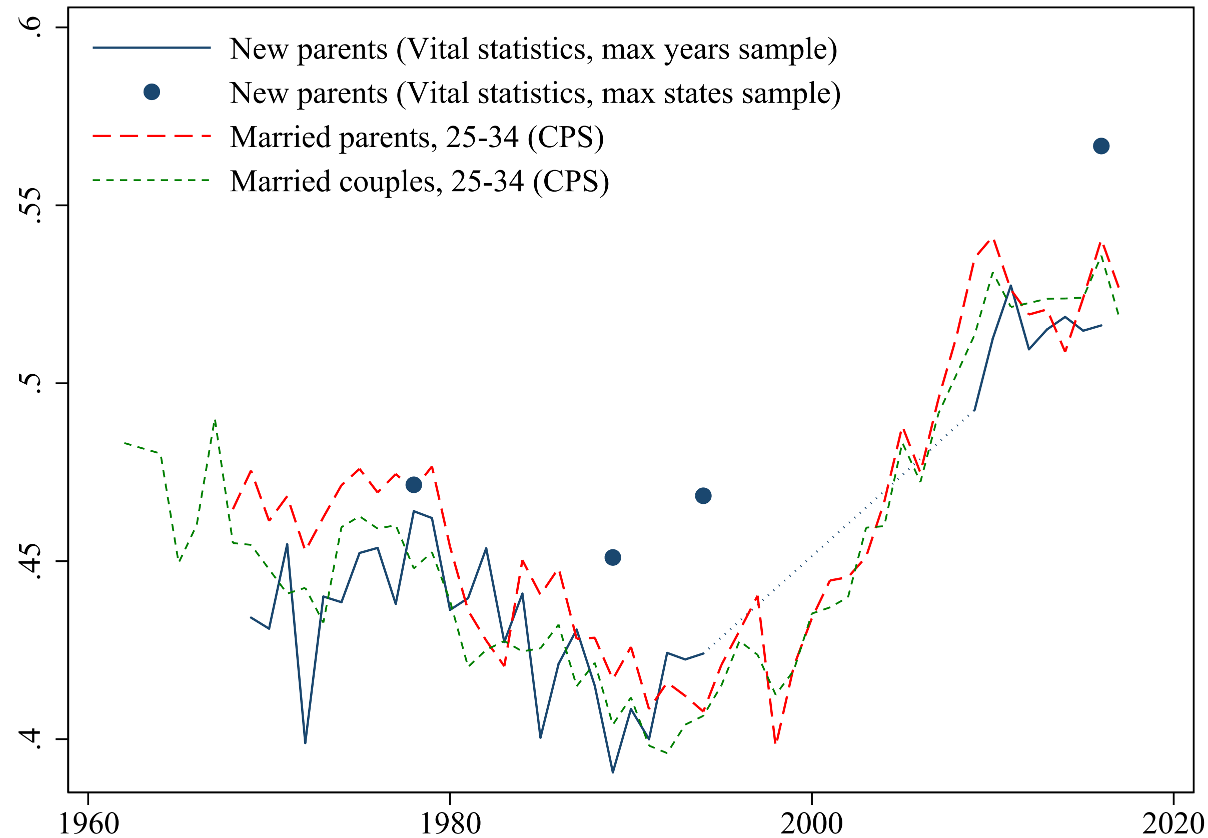
Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. [talk about different education categorizations] Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High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

Figure 11: Mechanics behind observed trends: Education correlation for observed matches vs. simulated matches



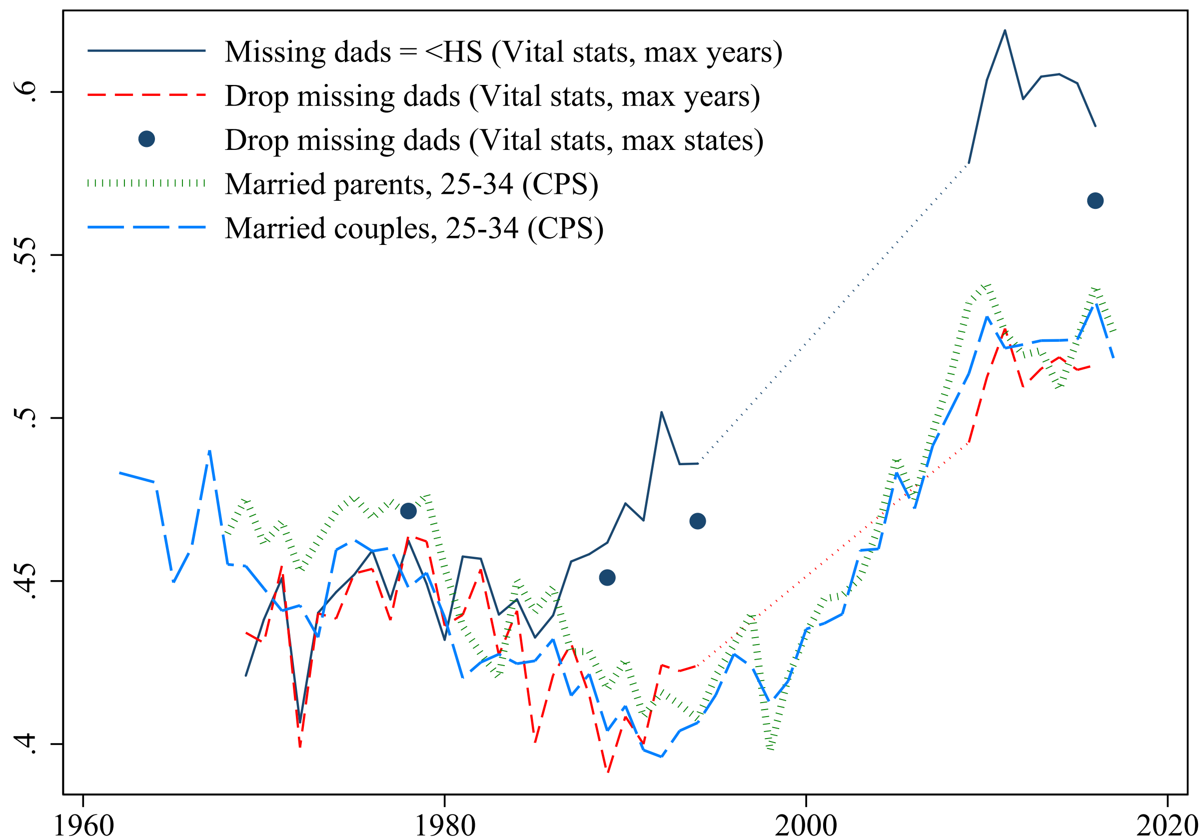
Source: Current Population Survey, ASEC/March supplements
Notes: Sample comprises all married couples where at least one spouse is between the ages of 26-60. Education is measured as follows: 1 = Less than high school (<12 years of schooling); 2 = High school graduate (12 years of schooling); 3 = Some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

Figure 12: Trends in educational assortative mating, 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 13: Trends in educational assortative mating, 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).

Tables

Table 1: Summary statistics, by sample

		Vital Statistics Births		
	(1)	(2)	(3)	(4)
	CPS	1% sample	Max years (1%)	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. See Section 6 for further discussion.

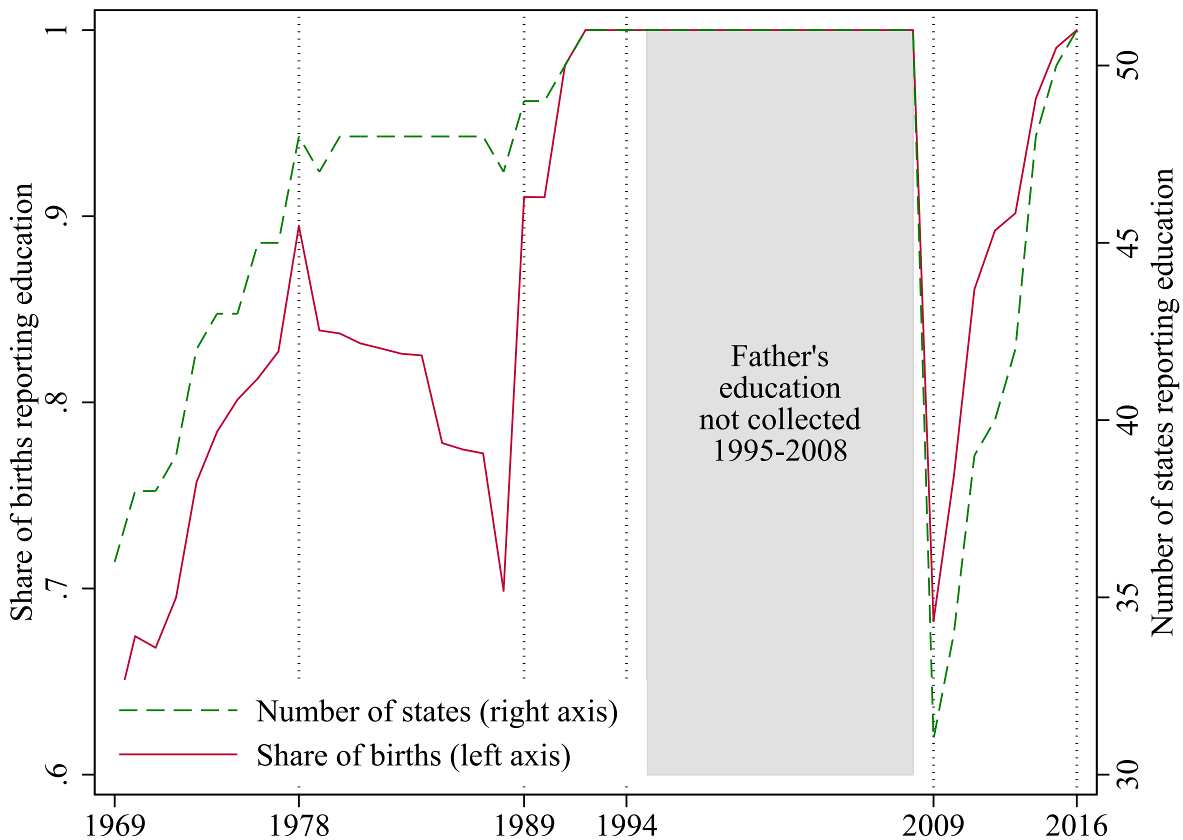
Table 2: Summary statistics of mothers and children by father's education

	Father's education				
	(1)	(2)	(3)	(4)	(5)
	Coll+	SC	HS	$<$ HS	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. See Section 6.2 for further discussion.

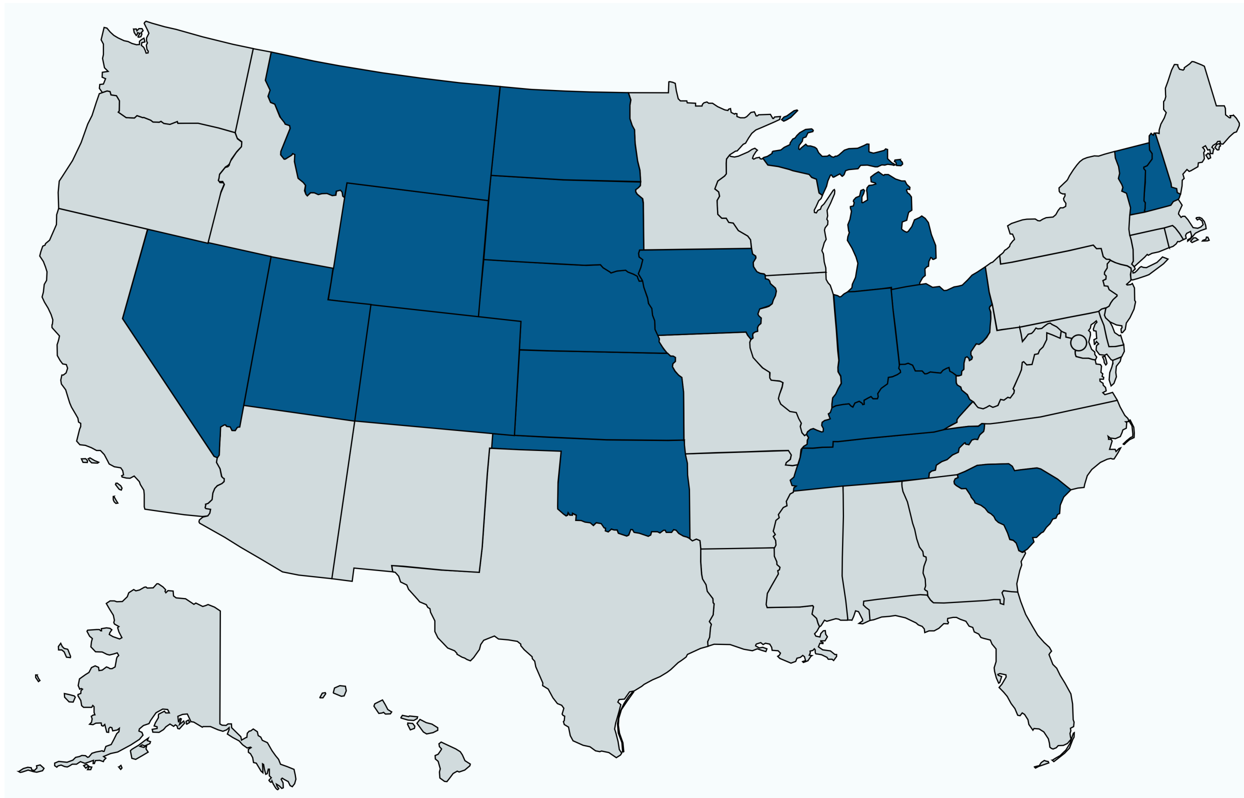
Appendix Figures

Figure A.1: 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 A.2: States in the max years sample

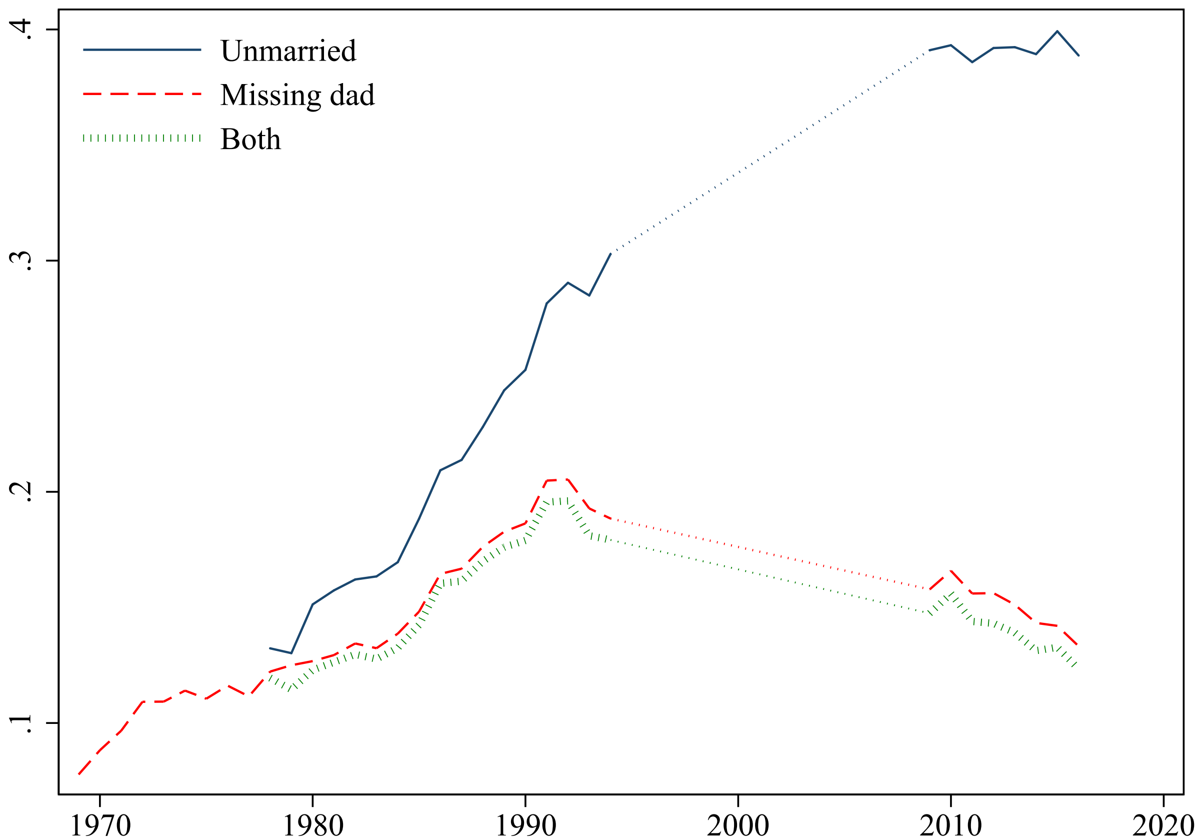


Notes: States shaded in blue consistenly report mother's and father's education between 1969 (the first year of available data) to present day. These states comprise the "max years" sample that I analyze in the Vital Statistics Natality dataset.

Figure A.3: Trends in educational assortative mating as measured by the Perfect-Random Normalization, simulation exercise vs. births samples, CPS

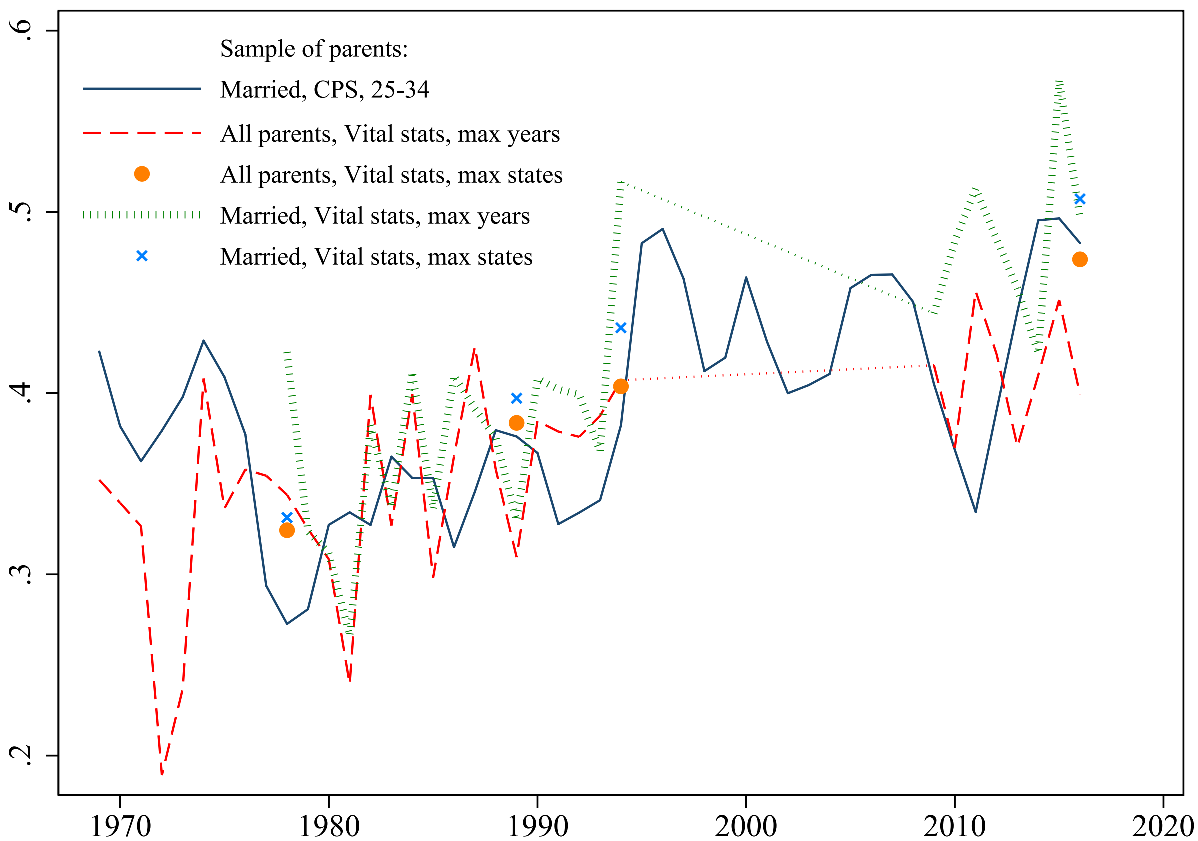
Source: Current Population Survey
Notes: Sample for dashed red line is all married couples (with spouse present) in the CPS where: (1) At least one spouse is aged 26-60 (the sample for the simulation exercise). Sample for solid blue line is all married couples (with spouse present) in the CPS where: (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 A.4: Share of births where mother is unmarried and/or where father's education is missing



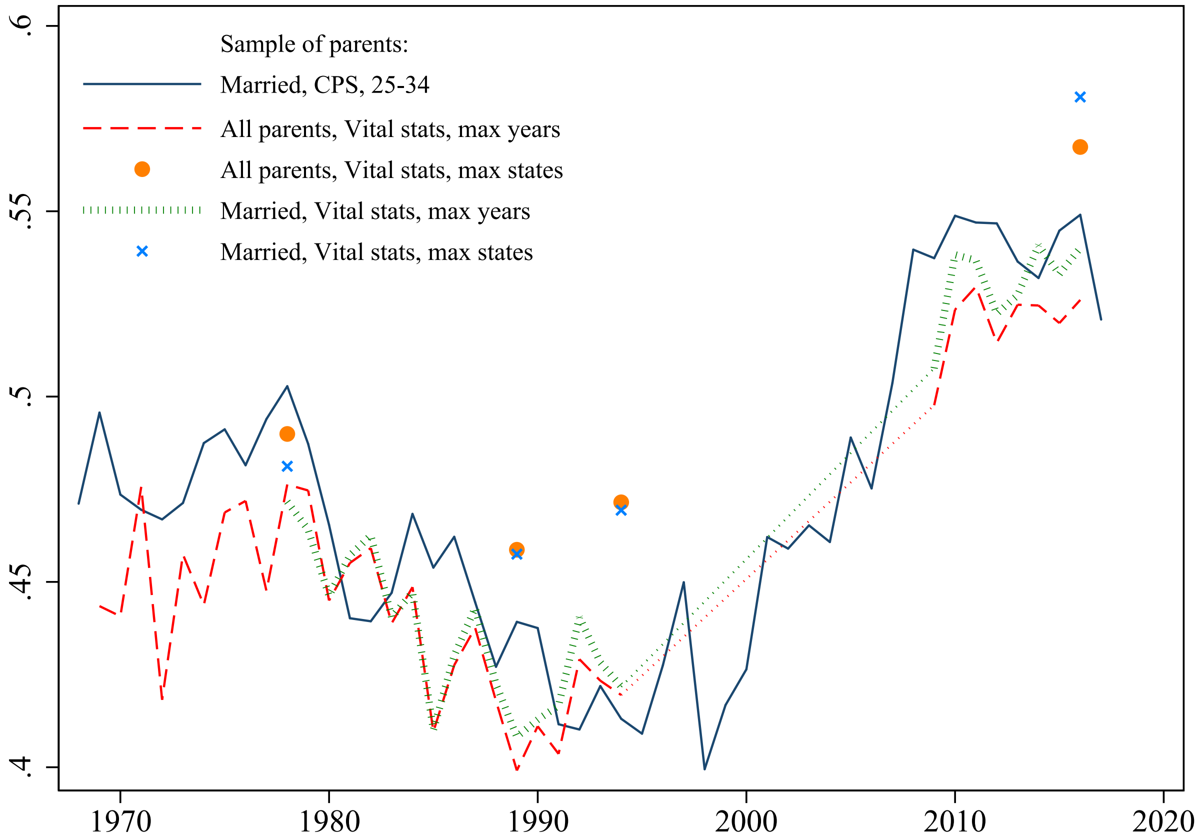
Source: Vital Statistics Natality, Max years 1% sample
Notes: Missing dad is defined as births where mother's education is reported but father's education is not. Sample is all births in states in the max years sample. See Section 6 for further discussion.

Figure A.5: Trends in educational assortative mating for couples where the woman is black



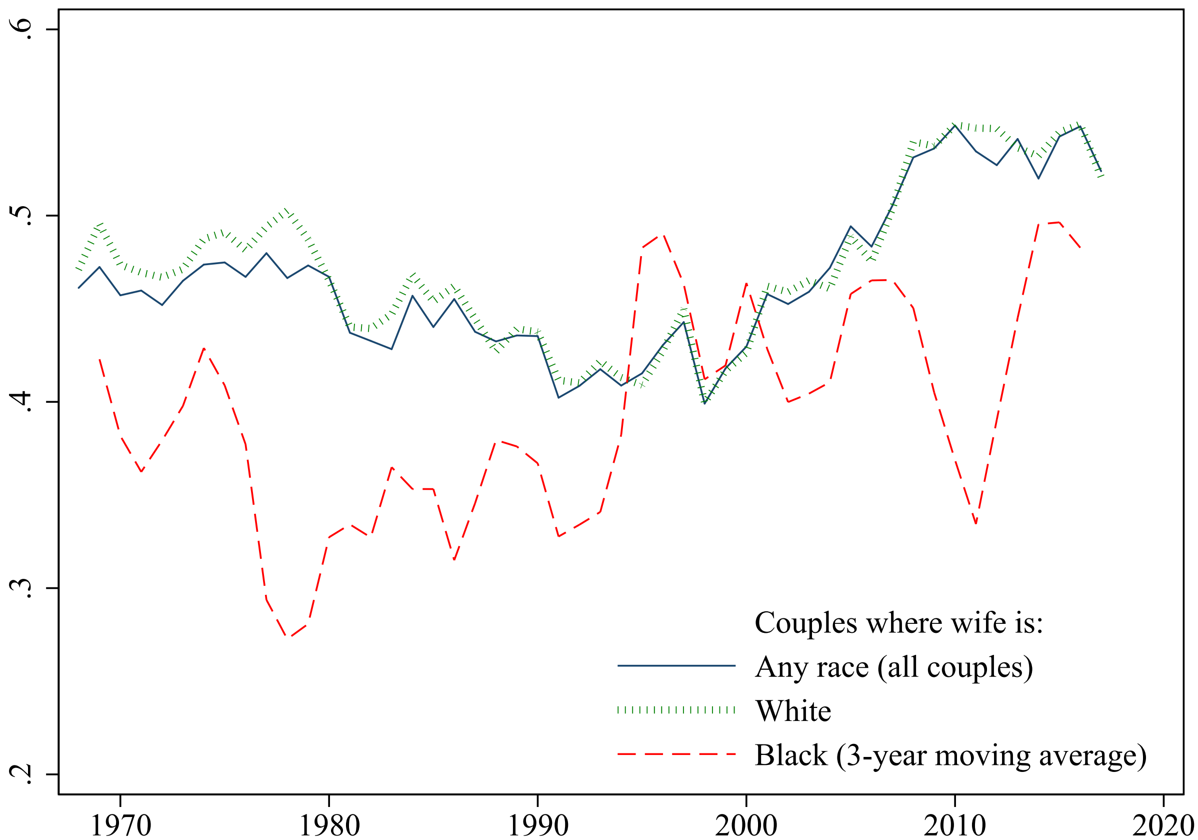
Source: Current Population Survey, Vital Statistics
Notes: For all samples, I restrict to couplles where the female partner is black. The CPS sample comprises all married couples where at least one spouse is between the ages of 25-34, at least one child under age 5 is present in the household. The vital 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. See Section 6 for further discussion.

Figure A.6: Trends in educational assortative mating for couples where the woman is white



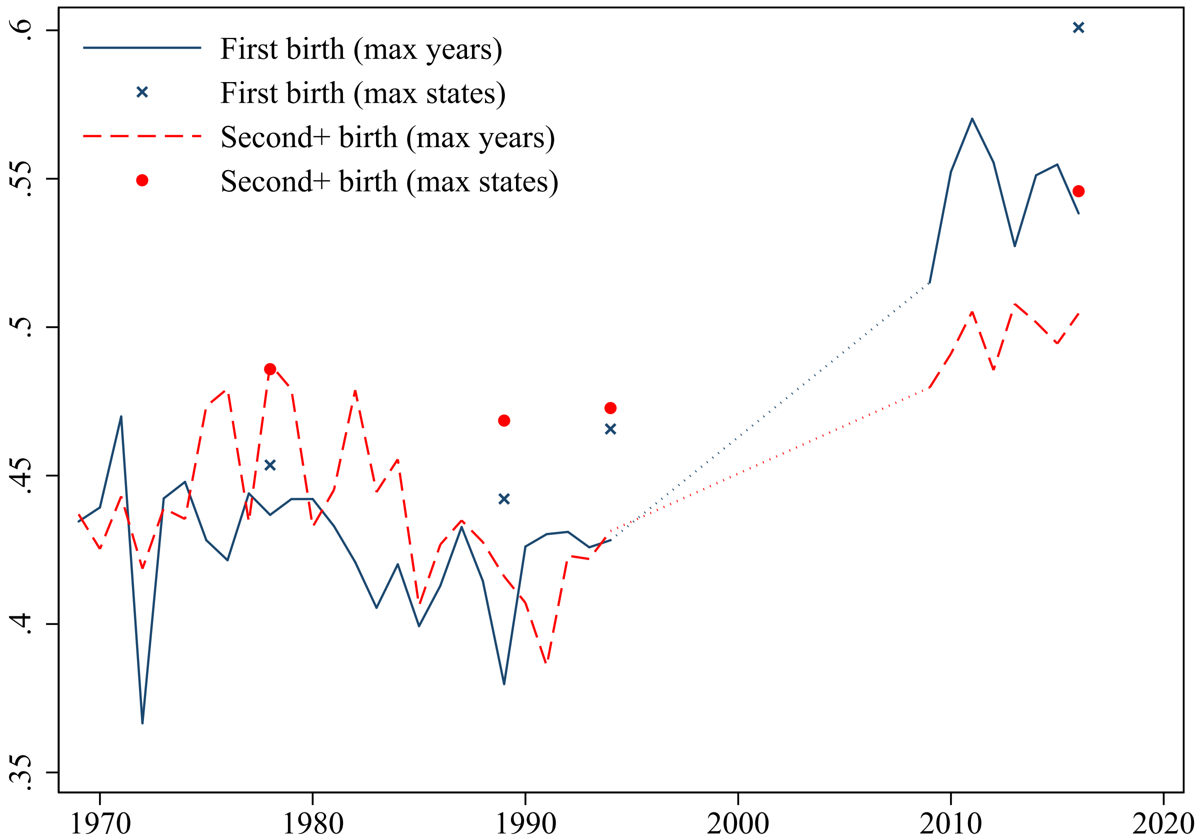
Source: Current Population Survey, Vital Statistics
Notes: For all samples, I restrict to couplles where the female partner is white. The CPS sample comprises all married couples where at least one spouse is between the ages of 25-34, at least one child under age 5 is present in the household. The vital 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). See Section 6 for further discussion.

Figure A.7: Trends in educational assortative mating 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 A.8: Trends in educational assortative mating 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. See Section 6 for further discussion.