Return to Education, Marriage Market, and Income Inequality

Mohammad Hoseini*

September 20, 2024

Click here for the latest version.

Abstract

Previous studies decomposing the growth of household income inequality based on marriage market outcomes, find negligible impact for the increase in assortative matching by education. We argue that the observed negligible effect is a consequence of the conditional independence assumption inherent in the decomposition exercise. Using a frictionless matching model with imperfectly transferable utility, we relax this assumption and account for the general equilibrium effect of return to education on marriage market outcomes. Estimation of the model using CPS data demonstrates that accounting for the monetary gains of marriage that drives assortative matching has a sizable impact on the growth of cross-sectional household inequality in the US. Between 1962 and 2023, this factor explains about 40 percent of the rise in Gini coefficient.

JEL classifications: I24, I26, J12

1 Introduction

Over the past century, income inequality has been on the rise in various regions of the world, and a large body of literature has explored plausible explanations for this trend. Starting from (Becker, 1973, 1974), one branch of the literature investigates the role of marriage market and in particular the impact of assortative matching (AM) on raising income inequality. However, even though in many countries, a significant increase in AM has been documented,[1] recent studies that decompose inequality find a negligible effect of the rise in AM in explaining the increase in inequality. For example, the prominent contribution by Eika, Mogstad, and Zafar (2019) finds that substituting the sorting pattern of 2010 with the pattern of 1962 explains less than 2 percent of the rise in the Gini coefficient of married couples between these two years in the US. They report similar results regarding the effect of changes in AM

---

*Tehran Institute for Advanced Studies, Khatam University. E-mail: m.hoseini@teias.institute.

[1] See Greenwood, Guner, Kocharkov, and Santos (2014, 2015); Chiappori, Costa Dias, and Meghir (2020b, 2021); Hryshko, Juhn, and McCue (2017); Breen and Salazar (2011) for the U.S., Chiappori, Costa-Dias, Crossman, and Meghir (2020a) for the UK, among others.

on changes in household inequality in Denmark, Germany, Norway, and the UK. Other studies on this subject also reach the same conclusion (Chiappori et al., 2020a; Dupuy and Weber, 2022).

In this paper, we argue that this negligible effect of changes inside marriage market on changes in income inequality primarily stems from the assumption of conditional independence in the standard inequality decomposition framework (Fortin, Lemieux, and Firpo, 2011). Under this assumption, the conditional income distribution for each type of couples remains fixed when the sorting pattern changes in the counterfactual scenario. Such an assumption overlooks the general equilibrium effects resulting from secular trends in the return to education on marriage market outcomes. As macroeconomic factors change the return to education, the economic gains associated with different marriages may experience disproportionate shifts across educational groups, thereby altering the incentive structure for marriage and marital sorting by education. In the standard counterfactual practice, it is assumed that while the economic gains of marriages is at its current level, households marry and sort in the same pattern as the base year.

Our primary objective is to assess the contribution of marriage market on cross-sectional income inequality by controlling for secular trends in the marginal population and income within the economy. To achieve this, we utilize an old statistical literature that demonstrates the representation of any matrix as two vectors of marginal distributions for rows and columns, along with a matrix of odds ratios indicating the association between rows and columns. In a population contingency table based on education, the marginals represent the population categorized by gender and education level, while the association matrix characterizes AM in the marriage market.

A key advantage of decomposing a couple's population table based on marginal distributions and row-column associations is that it allows us to disentangle the effects of changes in matching patterns from changes in the return to education on income inequality. However, these two factors do not evolve independently. To establish the link between macroeconomic trends and the marriage market, we apply the frictionless matching model with imperfectly transferable utility developed by Galichon, Kominers, and Weber (2019) (henceforth GKW). This model connects household formation to the allocation of power within households, both of which are determined in the marriage market equilibrium.

The main result of the theoretical model is that marriage market outcomes, including marriage rates and AM, are functions of the population ratios of singles and the marriage surplus, which is defined as the joint gain from marriage minus the sum of the gains each individual would have if they remained single. The surplus consists of two components: one coming from non-monetary gains, which is independent of the return to education, and the other related to monetary gains from marriage, which depends on the return to education and its secular trends in the economy. Both components can be identified using contingency tables of population and average income, along with an assumption about the income-sharing rule within each couple type. The theoretical model allows us to construct various counterfactual scenarios of income inequality by fixing either the non-monetary or monetary gains from marriage.

We estimate counterfactual income inequality for the US using Current Population Survey (CPS) data from 1962 to 2023. Consistent with the existing literature, we observe an upward trend in AM over this period. When replicating the standard decomposition exercise for the US, we find that AM has a negligible impact on income inequality. However, our counterfactual analysis reveals a significant reduction in income inequality when the monetary gains from marriage are fixed at their 1962 levels. This scenario suggests a reduction of 4 Gini points for both married couples and all households, accounting for approximately 40 percent of the overall rise in income inequality between 1962 and 2023. While the pattern of non-monetary gains from marriage also changed during the study period, its contribution to explaining shifts in cross-sectional household inequality is minimal.

This paper contributes to the literature on inequality from a household economics perspective. Since Becker (1973, 1974), AM has been a focal point in studies exploring household income inequality within the marriage market. When individuals with similar levels of education or skills form partnerships, they collectively possess more divergent earning potential, leading to more income inequality, compared to when partners have different levels of human capital. Although theory suggests strong link between AM and inequality (e.g. Fernández and Rogerson, 2001), the impact is found to be negligible in empirical studies (Kremer, 1997; Greenwood et al., 2015; Eika et al., 2019; Chiappori et al., 2020a; Dupuy and Weber, 2022). A missing component in these empirical analyses is the absence of general equilibrium effects between AM and the return to education, despite theoretical propositions that AM is positively related to the market return to human capital (Fernandez, Guner, and Knowles, 2005; Chiappori, Salanié, and Weiss, 2017). In particular, Chiappori et al. (2017) show that AM, measured by supermodularity of the surplus, is increasing in return to education.

Standard decomposition methods used in empirical studies typically impose the identification constraint that income distribution, conditional on education, is independent of changes in AM in counterfactual scenarios (Fortin et al., 2011). This paper challenges that assumption and seeks to "open the black box" of AM. Our first contribution is to link AM and return to education by decomposing AM into non-monetary and monetary components. The second contribution is to endogenize the decision to marry within the model, connecting it to the return to education. While the first contribution focuses on the intensive margin, examining the impact of return to education on the marriage market, the second addresses the extensive margin by exploring how return to education influences the decision to marry.

The paper also adds to the literature on the role of human capital in explaining inequality in the US (e.g. Goldin and Katz, 2009; Autor, 2014). By connecting return to education and partner choice, we provide new evidence on the channel that human capital affects inequality through the marriage market. In contrast to previous findings decomposing inequality, we demonstrate that marital sorting has a non-negligible size in the analysis of US inequality.

The rest of the paper is organized as follows: Section 2 describes the data and presents overall trends. Section 3 outlines the measurement of AM based on the association of row and columns of matching tables and presents the trend of AM in population and income. Section 4 reviews the standard decomposition practice and analyzes the challenges associated with relaxing the assumption of invariance in the conditional income distribution within the counterfactual scenario. Section 5 develops the matching model and discusses its identification and section 6 describes estimation and the procedure to build various counterfactual scenarios. Section 7 presents the results and finally, section 8 concludes.

2 Data and Overall Trends

We use the US Current Population Survey (CPS) for 1962-2023 which is the common dataset to study income inequality in the US. We consider marriage as a monogamous relationship, meaning there is an equal number of men and women matched with a partner at each point in time. Each year, the sample is restricted to either single individuals aged 26 to 60, excluding widowed individuals,[2] or married and cohabiting couples where at least one partner is between 26 and 60 years old. Information on cohabitation is unavailable in the CPS prior to 1995, so for those earlier years, we cannot distinguish cohabiting couples from singles. As a result, there is a slight jump in the number of couples observed in the CPS starting in 1995. Still, because cohabitation was rare in 1960s lack of cohabitation data is not a big concern for our counterfactual exercises. We exclude all single individuals and couples with missing data on age, education, or income.

Regarding educational classification, we assign individuals into five categories:

1. *Dropouts (D)*: those who have less than 12 years of education or have no high school qualification
2. *High school (HS)*: those who finished high school
3. *Some college (SC)*: those who attend 1 to 3 years of college, including associate's degree
4. *Bachelor's (BA)*: those who have bachelor's degree
5. *Graduate (G)*: those who have higher education than bachelor's degree

We begin by examining the trends in the distribution of education by gender. Figure 1 shows the changes in educational attainment across genders. Between 1960 and 1990, there is a significant decline in the proportion of individuals who did not complete high school, accompanied by an increase in the share of those with college degrees or higher, for both men and women. After 2000, these population shares remained relatively stable.

The second factor to consider is the return to education for both men and women, measured by the average income of all individuals in the respective group. Figure 2 illustrates the trend of average income by gender and education in 1999 dollars by considering zero income for non-participants. The dashed lines

---

[2] We exclude widowed individuals because their single status is unintentional. However, including them has a negligible impact on our main findings and mainly affects marriage rates, particularly for women, who are more likely to be widowed. This is also the case if we exclude divorced and separated individuals from the sample of singles.

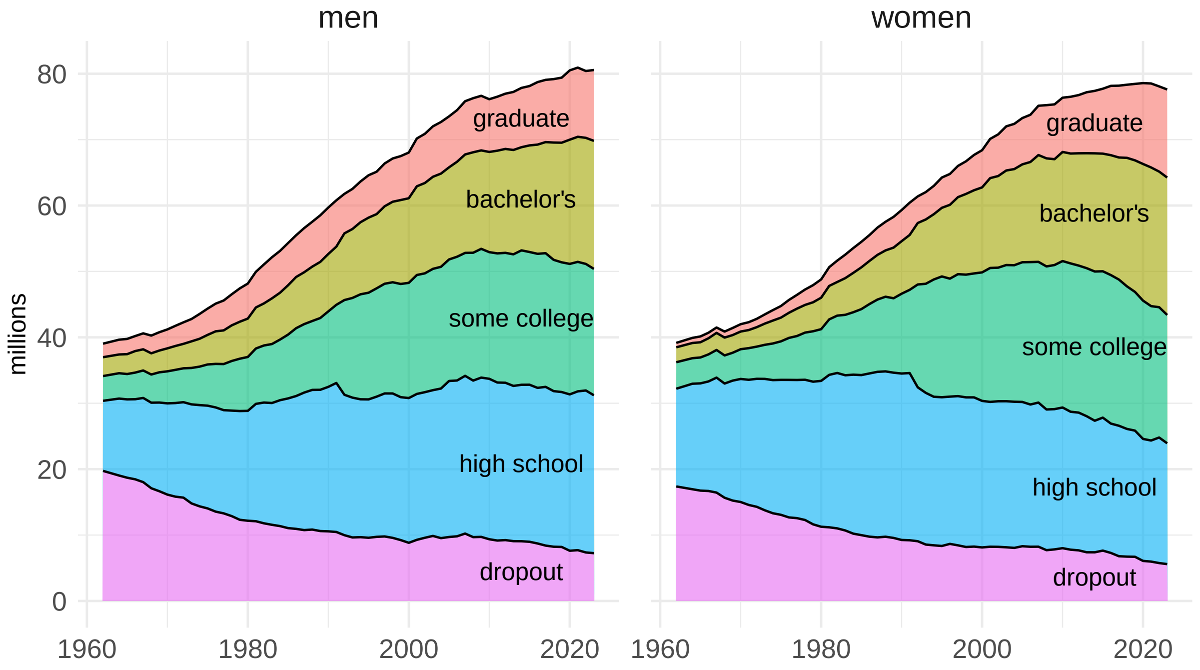


Figure 1: Trend of total population by education and gender in US. Data source: CPS, individuals between 26 and 60 years old.

show the average trend across all groups and the other lines show the trend among different educational groups. Concerning the average income level, we observe a slight increasing trend with cycles for men and a more pronounced upward trend without cycles for women. This gender heterogeneity can be attributed primarily to the increasing trend of participation among women, while men exhibited consistently high labor force participation during this period. Regarding the gap between different education levels, we observe a clear divergence for men over time. The divergence also exists for women to a lesser extent, and among those with education above high school level, the gap exhibits a U-shaped pattern.

Figure 3 illustrates the trend in marriage rates by gender and education. We observe a sharp decline in marriage rates for both men and women over the study period. For men, the rates are similar across education levels at the start, but the subsequent decline is inversely related to education level, with those holding graduate degrees having the highest marriage rates in recent years. A similar trend is seen for women, except for those with graduate education, who initially had significantly lower marriage rates, followed by an upward trend over time.

Given I educational categories for men and J for women, a matching table is a $(I+1) \times (J+1)$ two-way contingency table for the population. The rows correspond to men with education levels $i \in 1, \dots, I$, and the columns correspond to women with education levels $j \in 1, \dots, J$. The table also includes the single population, with a dummy partner index of 0. In this context, $N_{00} = \emptyset$, N_{i0} (N_{0j}) represents the population of single men (women), and for all $i, j > 0$, N_{ij} denotes the population of couples in which the man has education level i and the woman has education level j. For simplicity, we use $\oplus$ and $+$ in the

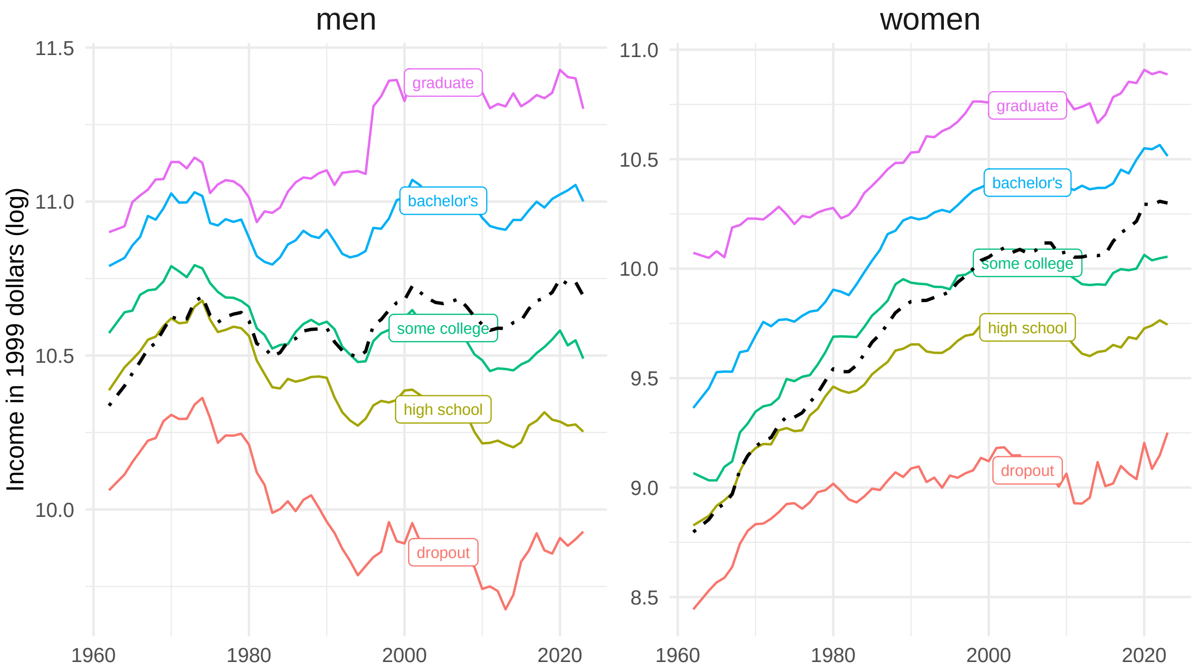


Figure 2: Trend of log average income by education for men and women in the US. For taking average, the income of non-participants are considered as zero. The dashed line show the average across the whole population. Income is adjusted by the 1999 price index. Data source: CPS, individuals between 26 and 60 years old.

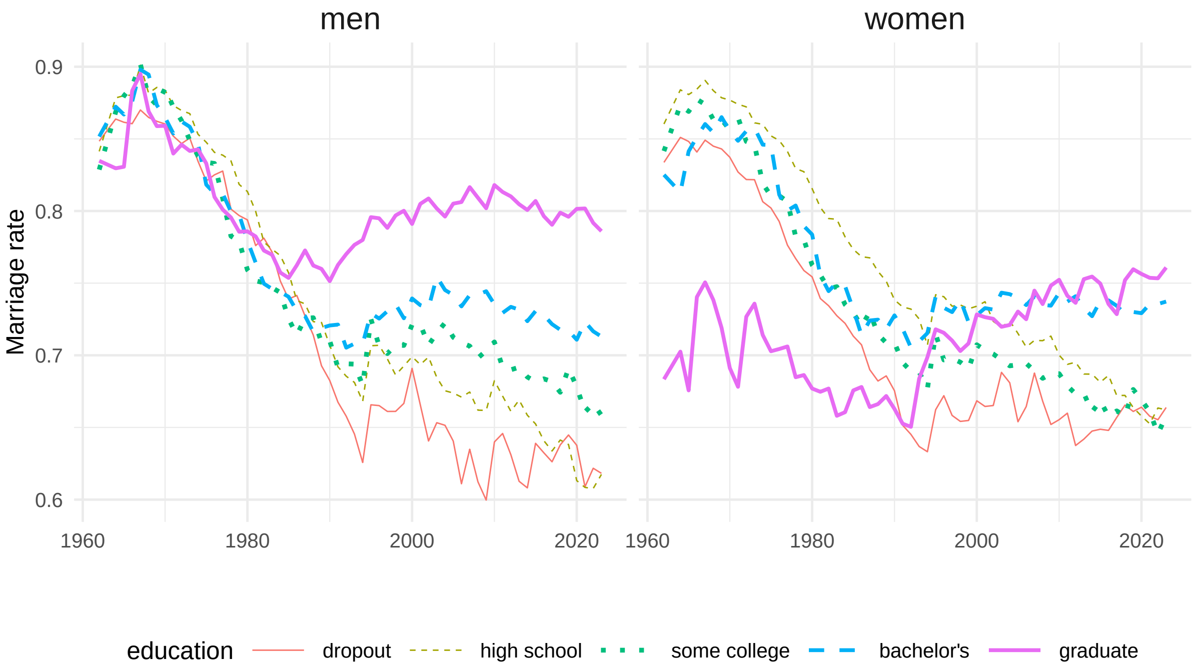


Figure 3: Trend of marriage rate by education for men and women in the US. Data source: CPS, individuals between 26 and 60 years old.

subscript to denote summation starting from 0 and 1, respectively. Thus, N_{i+} represents the population of married men with education level i, and $N_{i\oplus} = N_{i0} + N_{i+}$ represents the total population of men with education level i. Similarly, N_{+j} and $N_{\oplus j}$ represent the populations of married women and all women with education level j, respectively.

Using this notation, the marriage rates for men with education i and women with education j are defined as $\mu_i = N_{i+}/N_{i\oplus}$ and $\omega_j = N_{+j}/N_{\oplus j}$, respectively. These two indices are extensive margin measures that capture participation in the marriage market by education and gender. In the next section, we define a measure for assortative matching (AM), which is an intensive margin index capturing spouse quality by education, conditional on marriage.

3 Assortative Matching by Education (AM)

Measuring assortative matching is challenging and there are a variety of indices to measure assortativeness in the marriage market in the literature. Chiappori et al. (2021) examine the properties of the different sorting indices for a 2×2 contingency table and among them the log odds ratio ($\ln \frac{N_{11}N_{22}}{N_{12}N_{21}}$) is preferable for two reasons: First, it is independent to changes in the marginal distribution of the populations; second, it has a useful structural interpretation from the frictionless marriage market models of Choo and Siow (2006). For 2×2 tables, a single odds ratio can summarize the association, but for bigger tables, it is not possible to summarize association by a single number with no loss of information. Therefore, assortativeness should primarily treated as a local property and its global indices can be locally invalid.

In general, a $I \times J$ matrix has $\binom{I}{2} \times \binom{J}{2}$ odds ratios, among which $(I-1) \times (J-1)$ can be chosen as independent. The set of independent odds ratios for a table is not unique, and different basic sets may be chosen based on the application. Two popular sets are the *nominal* odds ratios, measured with respect to either the first or last group, and the *local* log odds ratios, measured for two adjacent groups.[3]

$$\text{nominal (first): } \frac{N_{11}\ N_{ij}}{N_{1j}\ N_{i1}}, \qquad \text{nominal (last): } \frac{N_{ij}\ N_{IJ}}{N_{iJ}\ N_{Ij}}, \qquad \text{local: } \frac{N_{i-1,j-1}\ N_{i,j}}{N_{i-1,j}\ N_{i,j-1}}, \qquad i,j > 1$$

Any of these sets comprises $(I-1) \times (J-1)$ elements that can be directly computed from the elements of another set. Here, to better illustrate AM, we present the set of log odds ratios benchmarked with the geometric average of the population, defined as

$$\rho_{ij} = \ln \frac{N_{ij}\ \bar{N}_{\times\times}}{\bar{N}_{i\times}\ \bar{N}_{\times j}} \tag{1}$$

where $\bar{N}_{i\times} = \prod_{j=1}^{J} N_{ij}^{1/J}$, $\bar{N}_{\times j} = \prod_{i=1}^{I} N_{ij}^{1/I}$ and $\bar{N}_{\times\times} = \prod_{i=1}^{I}\prod_{j=1}^{J} N_{ij}^{1/(IJ)}$ are the geometric means within j, i, and both, respectively. Note that this definition has a nice feature for illustration because $\sum_{i=1}^{I} \rho_{ij} = \sum_{j=1}^{J} \rho_{ij} = 0$. In other words, when computed for all $i, j > 0$, there is a redundant element

---

[3] See section 2.2.5 of Kateri (2014) for other common sets of odds ratios used in contingency table analysis.

in each row and column of the matrix ρ_{ij} such that the sum of all elements of a row or a column is zero. Moreover, the demeaned version of any basic log odds ratio set benchmarked by row i' and column j' is the geometric average set such that

$$\rho_{ij} = \text{LOR}_{ij,i'j'} - \frac{1}{J}\sum_{j=1}^{J}\text{LOR}_{ij,i'j'} - \frac{1}{I}\sum_{i=1}^{I}\text{LOR}_{ij,i'j'} + \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\text{LOR}_{ij,i'j'}$$

Figure 4 illustrates the trend of ρ_{ij}s overtime. We observe that the average levels are consistent with positive assortative matching, where the diagonal elements are significantly positive and the anti-diagonal elements are significantly negative. Furthermore, the trends of the elements indicate a movement toward increased AM over time, as evidenced by the majority of cases where the absolute values are rising. Appendix Figure 13 depicts the value of the log odds ratios of population at ten-year intervals from 1962 to 2022. This pattern also suggests a prevailing increase in assortative matching by education over time.

3.1 Aggregating AM indices

The above analysis show that AM is local properties, and for a $I \times J$ table, at least $(I-1)\times(J-1)$ odds ratios are needed for full characterization of AM. In this regard, any aggregation of AM elements involves information loss and is sensitive to the method. An important consideration in aggregation is the preservation of the attractive property of independence from marginal distribution that odds ratios possess. In this regard, a fixed weight must be applied across different points in time or space to achieve a marginal-free aggregate index (Hoseini, 2023).

A well-known aggregator of odds ratios for two-way tables is the metric of association proposed by Altham (1970) which is defined as:

$$\frac{1}{IJ}\sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{I}\sum_{l=1}^{J}\left(\ln\frac{N_{ij}\;N_{kl}}{N_{il}\;N_{kj}}\right)^2} \tag{2}$$

Altham's metric computes the root sum of squares of all $\binom{I}{2}\times\binom{J}{2}$ log odds ratios of a contingency table, with its value reflecting the degree of association between rows and columns. In the case of random matching, Altham's metric is zero, and higher values for a given table size indicate a greater distance from random matching. However, Altham's metric focuses on the absolute value of association and does not indicate whether the association is positive or negative. To address this limitation, we compute aggregate indices using the weighted average of the sets of odds ratios defined in (1):

$$\rho = \sum_{i=1}^{I}\sum_{j=1}^{J}\rho_{ij}\sum_{t=1}^{T}\frac{N_{ij}^{t}\;\bar{N}_{i\times}^{t}\;\bar{N}_{\times j}^{t}}{T\sum_{k}\sum_{l}N_{kl}^{t}\;\bar{N}_{k\times}^{t}\;\bar{N}_{\times l}^{t}} \tag{3}$$

To maintain the marginal-free property for the aggregate index, necessary for trend analysis over time,

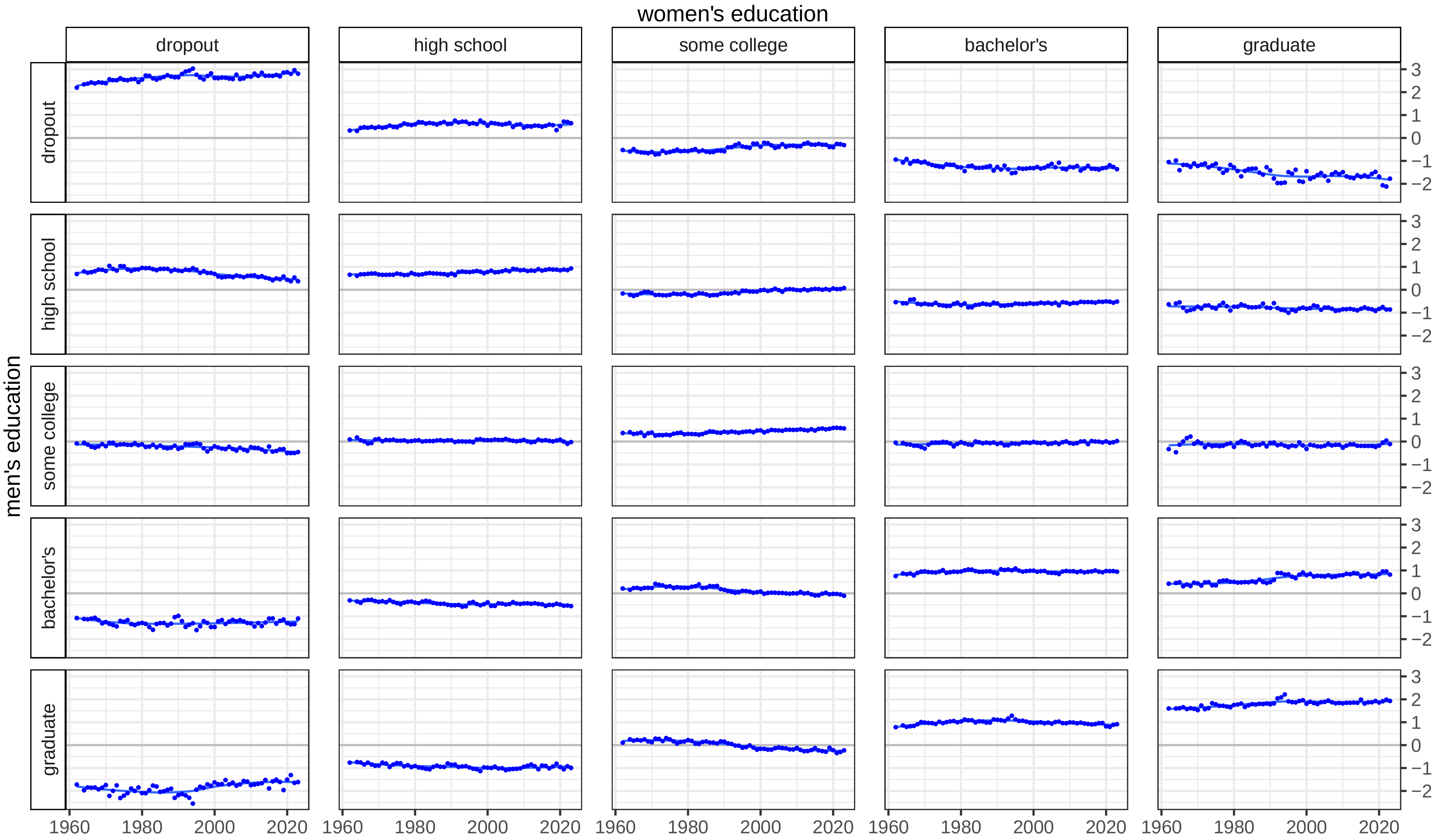


Figure 4: Assortative matching in population measured by log of geometric mean of odds ratios as $\ln \frac{N_{ij} N_{\times\times}}{N_{i\times} N_{\times j}}$.

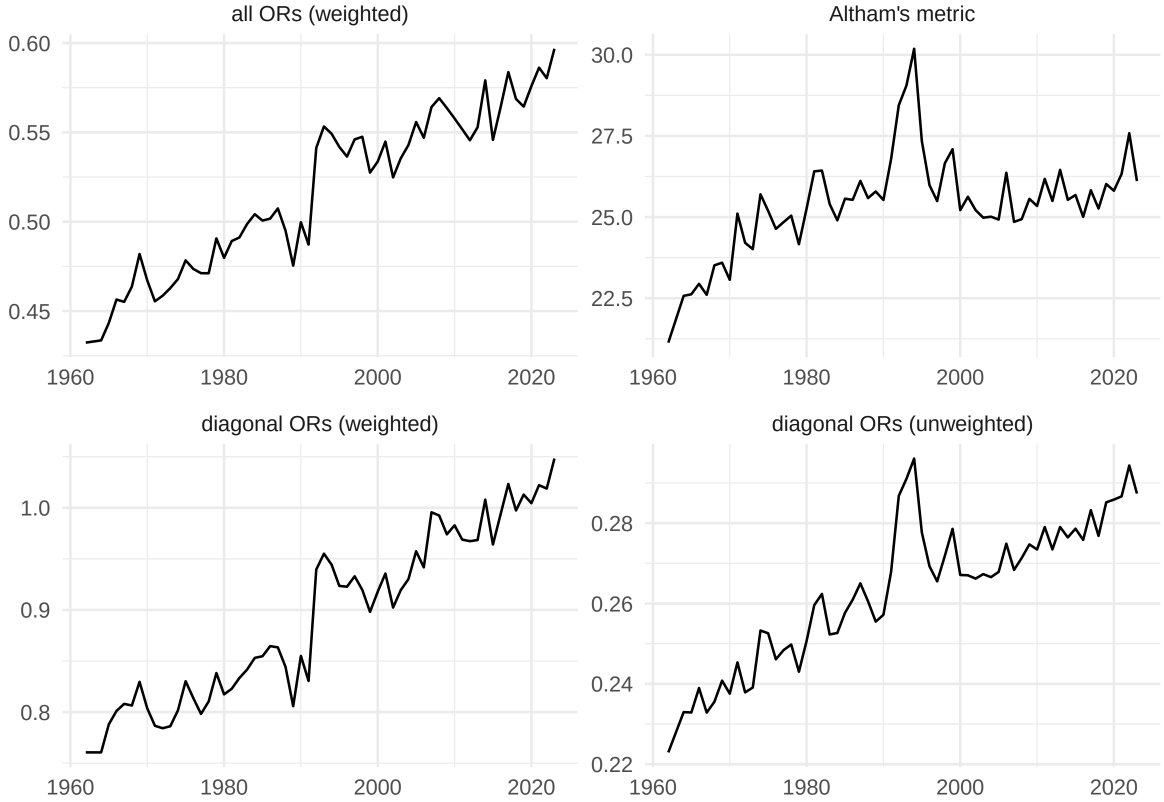


Figure 5: Different aggregate measures of assortativeness using log odds ratios.

we weight ρ_{ij} by the average of weights over all years. Since the AM property mainly manifests itself on diagonal elements, one can also compute the weighted and unweighted averages of diagonal elements as two additional aggregate indices. Various other aggregate indices exist in the literature (for a summary, see Figure 5 of Eika et al. (2019)), but they are not independent of changes in marginal distributions over time, so we do not consider them here.

Figure 5 illustrates the trends of different aggregate measures of AM. In the top right, Altham's metric shows an increasing trend between 1960 and 1980, leveling out thereafter. In the top left, the weighted average index, as defined in 3, indicates that AM is rising over the period of study. The bottom right and bottom left graphs show the unweighted and weighted average of the diagonal elements in (3), both of which display an increasing trend of AM.[4] Hence, we can conclude an increasing trend of AM in the US marriage market in the period of study.

3.2 Decomposing matching table by AM, marriage rates, and populations

The below proposition shows that a marriage contingency table can be characterized by AM matrix, marriage rates, and the marginal distribution vectors of the population. Intuitively, given that in one-to-one matching $\sum_i N_{i+} = \sum_j N_{+j}$, the marriage rate combined with marginal distributions provide

---

[4] Note that the unweighted average of log odds ratios (1) is zero.

$I+J-1$ independent equations. To determine the population of each couple type, we require $(I-1)(J-1)$ additional equations in the form of odds ratios. Proposition 1 affirms that a solution for such a system of equations always exists. This type of table decomposition serves as a valuable tool for disentangling the association between rows and columns from the marginal distribution of rows and columns. In our application, this implies the ability to separate the change in overall educational composition (measured by its marginal distribution by gender) from the marriage rates and the assortative matching between the two populations (measured by a basic set of odds ratios).

Proposition 1. *An* $(I+1) \times (J+1)$ *marriage contingency table is characterized by these components and vice versa*

- *two educational distribution vectors* $(N_{i\oplus})$ *and* $(N_{\oplus j})$*,*
- *two marriage rate vectors* μ_i *and* ω_j*, such that* $\sum_{i=1}^{I} \mu_i N_{i\oplus} = \sum_{j=1}^{J} \omega_j N_{\oplus j}$*, and*
- *an* $(I-1) \times (J-1)$ *educational assortative matching matrix including any basic set of odds ratios.*

The proof is based on Sinkhorn's theorem that asserts the existence and uniqueness of a contingency table based on its odds ratio set and its marginal sums. While decomposing the table to its components is straightforward, the characterization of a table from the component involves solving an system of non-linear equations at a size equal to the unknown elements of the contingency table. The common algorithm to find the elements is Iterative Proportional Fitting (IPF) that dates back to Stephan (1942).

Proposition 1 provides a great tool to investigate marriage market outcomes independent of changes in population supplies. It asserts that one can build a marriage table with elements $\left(N_{ij}\right)$ from marginal population vectors $\left(N_{i\oplus}\right)$, $\left(N_{\oplus j}\right)$, marriage rate vectors $\left(\mu_i\right)$, $\left(\omega_j\right)$, and the AM matrix $\left(\rho_{ij}\right)$. This means that we can make counterfactual exercise by fixing any of these component at a benchmark year and find the equilibrium matching table.

4 Decomposition of Income Inequality

DiNardo, Fortin, and Lemieux (1996) introduce the standard decomposition practice to assess the contribution of different factors in income inequality. Let $F_{\mathcal{Y}|\mathcal{X}}(y|x,t)$ represent the conditional distribution of income by population group x. From the law of total probability, the income distribution at time t becomes:

$$F_{\mathcal{Y}}(y|\, t) = \int F_{\mathcal{Y}|\mathcal{X}}(y|\, x, t)\, dF_{\mathcal{X}}(x|\, t)$$

In a scenario in which the distribution of population is as in t_x, DiNardo et al. (1996) build the counterfactual income distribution as

$$\widehat{F}_{\mathcal{Y}}(y|\, t) = \int F_{\mathcal{Y}|\mathcal{X}}(y|\, x, t)\, \Psi(x|t, t_x) dF_{\mathcal{X}}(x|\, t), \qquad \Psi\left(x|\, t, t_x\right) = \frac{d\hat{F}_{\mathcal{X}}(x|\, t_x)}{dF_{\mathcal{X}}(x|\, t)}$$

where $\Psi\big(x|, t, t_x\big)$ is the reweighing function of the samples.

In our application, the population distribution N_{ij} is characterized by the four components described in Proposition 1. Following the approach proposed by DiNardo et al. (1996), we can construct the counterfactual inequality at time t when the educational distribution, marriage rate, and AM are at the levels of t_N, t_M and t_A, respectively, as:

$$\widehat{F}_{\mathcal{Y}}(y|\,t) = \sum_{i=1}^{I}\sum_{j=1}^{J} F_{\mathcal{Y}|\mathcal{I},\mathcal{J}}(y|\,i,j,t)\,\widehat{N}_{ij}\big(t_N, t_M, t_A\big)$$

Here, $F_{\mathcal{Y}|\mathcal{I},\mathcal{J}}(y|\,i,j,t)$ is the conditional income distribution for couples with education i and j, and $\widehat{N}_{ij}\big(t_N, t_M, t_A\big)$ is the counterfactual population when marginal population vectors are measured at alternative times. In the decomposition practice, usually one factor is benchmarked at the base year, while others vary over time. Then the change in the trend of inequality reflects the contribution of that factor in overall changes in inequality.

The decomposition method outlined above is applied in Eika et al. (2019) using the same dataset as ours, and we replicate their findings in Figure 6.[5] We observe that while fixing educational distribution as in 1962 significantly increases inequality, fixing AM as in 1962 has a negligible impact on inequality in the subsequent years. Despite different measures of AM suggesting an increasing trend, the counterfactual trend for constant AM seems surprising. One reason for this counter-intuitive result could be the assumption of invariant conditional distribution of income over time which assumes $F_{\mathcal{Y}|\mathcal{X}}(y|\,x,t)$ is fixed in the original and counterfactual scenarios.

As argued by Fortin et al. (2011), the conditional independence (or ignorability) assumption neglects the broader impacts arising from long-term trends in the returns to education on AM. Essentially, it supposes that, while the pecuniary gains of marriages, which depends on the return to education, is changing over time, households sort in the same pattern as the base year. However, during periods when macroeconomic factors significantly influence average income by education, the financial gains of marrying a partner with different human capital undergo uneven changes. Consequently, the absence of a connection between average income and AM fails to capture variations in the incentive structure for marital sorting resulting from macroeconomic factors.

Hence, we seek to to relax the assumption of exogenous conditional income distribution to changes in population distribution by allowing for the adjustment in the conditional income distribution in our counterfactual experiments. Formally, for a couple (m, w) in the sample belonging to education groups (i, j), we assume that the counterfactual income becomes

$$\widehat{y}_{mw} = \frac{\widehat{Y}_{ij}}{Y_{ij}}\,y_{mw} \tag{4}$$

---

[5] Our estimated Gini coefficient differ slightly from Eika et al. (2019), mainly because they used an older correction for top income coding of CPS based on Larrimore et al. (2008). More recently, IPUMS provides the corrected top income coding using an updated method proposed by Census Bureau. For more details, see: `https://cps.ipums.org/cps/topcodes_tables.shtml`

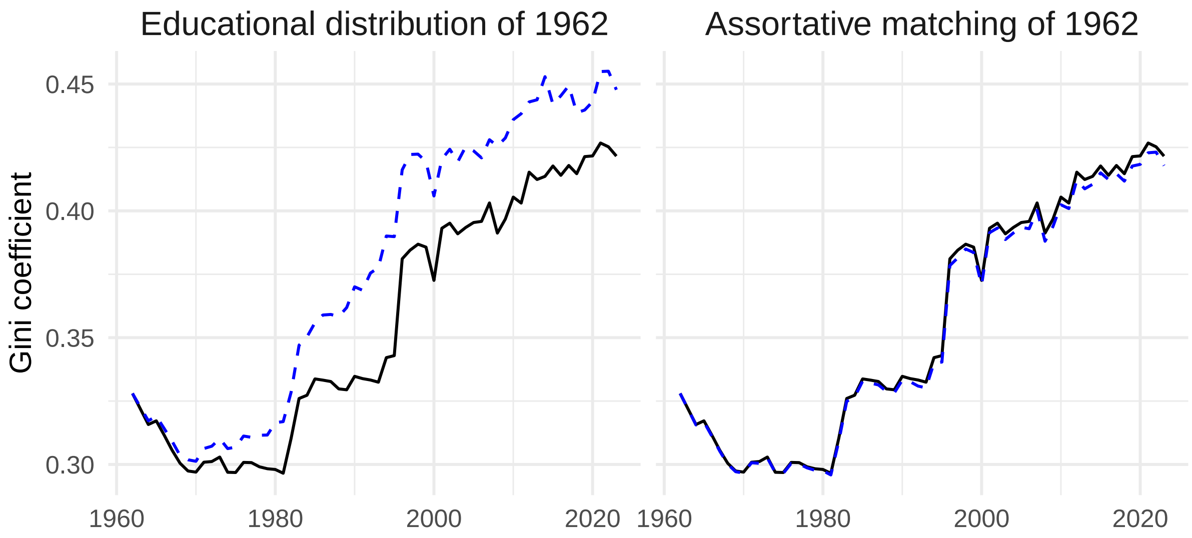


Figure 6: Counter-factual analysis with fixed conditional income distribution of couples. The solid lines are the real Gini coefficient and the dashed lines are the counterfactual estimations. The sample is married couples between 26 and 60. This Figure is replicated version of the same finding in Eika et al. (2019).

where $\widehat{Y}_{ij}$ is the adjusted average income of couples in educational groups ij after accounting for the impact of return to education, population change, etc.We then can simply compute a counterfactual inequality index like Gini coefficient by rewighting the sample multiplier for household mw using $\widehat{N}_{ij}/N_{ij}$ and considering $\widehat{y}_{mw}$ as their income.

To establish a connection between changes in AM and variations in the conditional income, in addition to estimate the matrix of $\widehat{N}_{ij}$, we need to construct matrices $\widehat{Y}_{ij}$ representing the predicted average income within each matched group in the counterfactual scenario. In general, the compositions of population and income in the economy are two important factors to determine income inequality. Both of these variables evolve based on exogenous factors outside the marriage market. Such secular trends in population and total income can be in their levels and composition by human capital. Since the income composition by education also plays a role in determining income inequality, it confounds assessing the impact of a change in AM on changes in household income inequality. Therefore, for a counterfactual exercise, we should control for these trends as much as possible. A big complication here is the dependency of AM to the changes in the average income by education. Since the gain from marriage is composed of economic and non-economic benefits, the state of the economy, namely the level and the composition of income by education, directly affect marriage gains via the economic benefits. These parameters are the outcome of equilibrium in the marriage market, and to find them we need a matching model that accounts for both economic and non-economic gains of marriage. In the next section, we provide a frictionless matching model with imperfectly transferable utility to determine the link between return to education and the marriage market outcomes.

5 Theoretical Model

In this section, we present the theoretical framework that we later use for counterfactual experiments. Since our goal is to characterize the relationship between the return to education and marriage market equilibrium, our model must account for both the matching decisions and the intrahousehold allocation of resources. To do so, we apply the matching framework with imperfectly transferable utility, as developed by GKW, which provides a proper tool for addressing such problems. This approach allows us to internalize the effect of the return to education, which influences household income, on matching decisions. At the household level, non-monetary gains from marriage are exogenous to household income and non-transferable. In contrast, monetary gains depend on the state of the economy, particularly the return to human capital, and these gains are imperfectly transferable between partners through consumption.

The main result of the theoretical model is that marriage market outcomes, including marriage rates and AM, are functions of the population ratios of singles and the marriage surplus, defined as the joint gain from marriage minus the sum of gains when both individuals remain single. Furthermore, the surplus consists of two components: one related to non-monetary gains and the other to monetary gains from marriage. Both components can be identified using the contingency tables of population and average income, along with an assumption about the sharing rule within each couple type.

5.1 Matching Model under Imperfectly Transferable Utility

The population is comprised from men and women, indexed by m and f, that may match and form couples. At the individual level, a matching is a dummy variable ν_{mf} which is one if m and f are matched and zero otherwise. We consider one-to-one matching such that each individual can match with at most one partner. This means that $\sum_f \nu_{mf} \leq 1$ and $\sum_m \nu_{mf} \leq 1$. Each matching ν generates payoffs u_m and v_f for man m and woman f, respectively. These payoffs determine feasibility and stability of the matching.

To characterize equilibrium matching when the utility is imperfectly transferable between the partners, GKW define $\mathcal{B}_{mf}$ as a proper bargaining set of feasible utilities (u_m, v_f) for m and f if it has three features: closed and nonempty, lower comprehensive, and bounded above.[6] A proper bargaining set has a corresponding distance-to-frontier function defined by

$$D_{mf}(u,v) = \min \left\{ z \in \mathbb{R} : (u - z, v - z) \in \mathcal{B}_{mf} \right\} \tag{5}$$

A matching is feasible when $D_{mf}(u,v) \leq 0$. Moreover, let u_{m0} and v_{0f} be the utilities of single men and women, respectively, then, a matching is stable if

- $\forall m, f: \; D_{mf}(u_m, v_f) \geq 0$ with equality when $\nu_{mf} = 1$,

---

[6] When utility is perfectly transferable, the set is the area below a line with slope -1.

- $u_m \geq u_{m0}$ with equality if $\sum_f \nu_{mf} = 0$ and $v_f \geq v_{0f}$ with equality if $\sum_m \nu_{mf} = 0$.

If $D_{mf}(u_m, v_f) < 0$ for a pair m and f, they would be better off by leaving their current status, matching together and sharing the extra attainable payoff.

5.2 Matching by categories

Suppose the population of men and women belong to a small number of categories and let $i \in \{1, \ldots, I\}$ and $j \in \{1, \ldots, J\}$ denote the types of men and women, respectively. For single individuals, we consider a dummy partner and denote it with a null category 0.

Assumption 1. *There exists families of non-vanishing distribution functions F_{α^j} and F_{β^i} such that*

- *if $m \in i$ and $f \in j$ are matched, for a proper bargaining set $\mathcal{B}_{ij}$, there exist $(U_m, V_f) \in \mathcal{B}_{ij}$, such that $u_m = U_m + \alpha^i_m$ and $v_f = V_f + \beta^j_f$,*
- *if m and f remain single, their utilities are $U_{i0} + \alpha^0_m$ and $V_{0j} + \beta^0_f$, respectively.*

where $\forall i \in \{0, \ldots, I\}, j \in \{0, \ldots, J\}, \alpha^j_m$ and β^i_f are random i.i.d vectors from F_{α^j} and F_{β^i}, respectively.

This assumption generalizes the concept of separability of unobservable heterogeneity in joint surplus, which is a key assumption in the literature on matching under transferable utility since Choo and Siow (2006). The non-vanishing property of the distribution in Assumption 1 ensures that all matches in the marriage contingency table have positive populations, preventing any zero cells. A slight modification in Assumption 1, compared to GKW, is the inclusion of systematic utilities for singles based on their category. In GKW and previous literature, U_{i0} and V_{0j} are benchmarked at zero, primarily because the discrete choice model can only identify differences in deterministic utilities within a type, requiring one category to be normalized. However, in what follows, we adopt a collective model where the utility of singles depends on their income, and thus we specify these systematic utilities as separate terms.

Under Assumption 1, the deterministic utilities U_m and V_f, which act as transfers, are allowed to vary within a type. However, GKW show that, with finite utilities,[7] this leads to an aggregate equilibrium where transfers depend only on the types of the match, meaning $U_m = U_{ij}$ and $V_f = V_{ij}$. The next proposition presents a simplified version of this result.

Proposition 2. *Under Assumptions 1 with bounded utilities, in a stable matching, there exists $2 \times I \times J$ numbers as U_{ij} and V_{ij} such that*

- *$D_{ij}(U_{ij}, V_{ij}) = 0$, where $D_{ij}(u, v)$ is the distance-to-frontier function of the bargaining set $\mathcal{B}_{ij}$,*
- *If $m \in i$ is matched with $f \in j$, their utilities are $u_m = U_{ij} + \alpha^j_m$ and $v_f = V_{ij} + \beta^i_f$.*

---

[7] The technical assumption in GKW is that the maximum utility any individual can obtain from matching with a partner of a given type is either always finite or always infinite.

A well-known assumption in discrete choice models that can substantially simplify the analysis is the use of the standard Gumbel distribution for all unobservable terms.

Assumption 2. *$\forall i, j$ $F_{\alpha^j}(\cdot)$ and $F_{\beta^i}(\cdot)$ are standard Gumbel (type-I extreme value) distribution.*

Proposition 3. *Under Assumptions 1 and 2:*

$$U_{ij} - U_{i0} = \ln \frac{N_{ij}}{N_{i0}}, \qquad V_{ij} - V_{0j} = \ln \frac{N_{ij}}{N_{0j}}, \qquad \ln N_{ij} = -D_{ij}(U_{i0} - \ln N_{i0},\ U_{0j} - \ln N_{0j})$$

Thus, when the utilities are additively separable and the unobserved heterogeneity has Gumbel distribution, number of matches in a couple type depends on the single's population and utilities in the respective categories.

For couple type ij, we define *marriage surplus* as the average surplus from marriage per partner

$$S_{ij} := \frac{1}{2}\big(U_{ij} + V_{ij} - U_{i0} - V_{0j}\big)$$

Proposition 3 implies that under Assumptions 1 to 2, the marriage surplus for couple ij is computed as

$$S_{ij} = \frac{1}{2} \ln \frac{N_{ij}^2}{N_{i0}N_{0j}} = -D_{ij}(U_{i0} - \frac{1}{2} \ln \frac{N_{i0}}{N_{0j}},\ V_{0j} + \frac{1}{2} \ln \frac{N_{i0}}{N_{0j}}) \tag{6}$$

The marriage surplus is a key factor in determining equilibrium in the marriage market. The following proposition illustrates the relationship between the marriage surplus and the marriage market indices.

Proposition 4. *Under Assumptions 1 and 2,*

$$\begin{aligned} \rho_{ij} &= S_{ij} - \frac{1}{I}\sum_{i=1}^{I} S_{ij} - \frac{1}{J}\sum_{j=1}^{J} S_{ij} + \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} S_{ij} \\ \mu_i &= \frac{1}{N_{i\oplus}} \sum_{j=1}^{J} \exp(S_{ij})\sqrt{N_{0j}N_{i0}} \\ \omega_j &= \frac{1}{N_{\oplus j}} \sum_{i=1}^{I} \exp(S_{ij})\sqrt{N_{i0}N_{0j}} \end{aligned}$$

This proposition illustrates the link between marriage surplus and the outcomes of the marriage market. From (6), we see that the surplus is determined by the distance-to-frontier function, characterized by intra-household decisions, and the utilities of singles and their population ratios. In general, the distance function does not have a closed-form representation, making the analysis complex. Therefore, in the following sections, we introduce additional structure to the household decision-making process to derive an analytical expression for the surplus.

5.3 Collective model for household decision

To model household behavior we employ a collective approach (Chiappori, 1992) in which the decisions are at the Pareto frontier. Let $u_i = \mathcal{U}(c_i)$ and $v_j = \mathcal{V}(c_j)$ be the utilities of a representative man in a category i and a woman in categories j as a function of their private consumption c_i and c_j. The budget constraint takes the form $c_i + c_j \leq Y_{ij}$, where Y_{ij} is the representative household income for private consumption which is observable in the data. Assuming that the utility functions are invertible, the budget constraint is a *proper bargaining set* by GKW's definition as follows

$$\mathcal{B}_{ij} = \Big\{(u,v) \in \mathbb{R}^2,\ \mathcal{U}_i^{-1}(u) + \mathcal{V}_j^{-1}(v) \leq Y_{ij}\Big\}$$

with a distance-to-frontier function defined by (5) as $D_{ij}(u,v)$.

In the collective framework, household solves

$$\max\ \lambda_{ij}\, u_i + (1-\lambda_{ij})\, v_j \quad \text{s.t.} \quad D_{ij}(u,v) \leq 0$$

where λ_{ij} is the Pareto weight associated with partner i which summarizes the allocation of power within the household (see Browning, Chiappori, and Weiss (2014), section 3.5). GKW show that when the bargaining set is smooth and convex, the Pareto weight is the derivative of the distance-to-frontier function with respect to its first argument

$$\lambda_{ij} = \partial_u D_{ij}(u,v) \tag{7}$$

This property integrates the allocation of power within the household with the matching process in the marriage market. This is particularly important for our analysis, as it allows us to model the impact of changes in the return to education on marriage market outcomes, both at the matching stage and through household decisions.

In the framework described above, since $D_{ij}(\cdot,,\cdot)$ specifies the Pareto frontier for households, it is generally a function of the total household income Y_{ij}. To characterize the distance-to-frontier function, we assume that, given a level of household income, $D_{ij}(\cdot,\cdot)$ takes on a parametric form that is a scaled version of a known distance-to-frontier function.

Assumption 3.

$$D_{ij}(u,v) = \gamma_{ij}\, d(\frac{u-a_{ij}}{\gamma_{ij}}, \frac{v-b_{ij}}{\gamma_{ij}}, Y_{ij})$$

where $d(\cdot,\cdot,y)$ is a known distance-to-frontier function which is decreasing in y.

Here, a_{ij} and b_{ij} align the means and γ_{ij} adjusts for the scale of the utilities.[8] Few classes of bargaining sets, including the ones explored in GKW, have closed-form distance functions. The most

---

[8] We need to multiply γ_{ij} to keep this property of distance-to-frontier function $D(x+u, x+v) = x + D(u,v)$

common form to model household decision in the previous literature, is the transferable utility (TU) with a distance function independent of income, such that $d(u, v) = \frac{1}{2}(u+v)$. Under TU, the distance function is $D_{ij}(u, v) = \frac{1}{2}(u + v - z_{ij})$, where $z_{ij} = a_{ij} + b_{ij}$ represents the joint gain from matching that can be freely transferred between spouses. This framework simplifies surplus estimation, as it requires only the identification of the joint gain z_{ij}, which can be determined from population observations in a single market. However, in our application, where we intend to link matching decisions with household income, TU is not the convenient model. In fact, TU assumes there is a (composite) good that serves as a constant exchange rate for transferring utility between partners. Equivalently, it imposes that the utility of both partners are linear with the same coefficient for the exchange good (see Chiappori and Gugl (2020) for more details), which is not a convenient assumption when transfer is made via private consumption while marriage creates other gains that cannot necessarily be cardinalized as private consumption.

5.4 Exponentially Transferable Utility

An alternative to TU for modeling household decision with imperfect transfer is the *Exponentially Transferable Utility* (ETU) as defined by GKW with

$$d(u, v; y) = \ln \frac{\exp(u) + \exp(v)}{y}$$

Under Assumption 3, the distance-to-frontier function of ETU is

$$D_{ij}(u, v) = \gamma_{ij} \ln \frac{\exp(\frac{u - a_{ij}}{\gamma_{ij}}) + \exp(\frac{v - b_{ij}}{\gamma_{ij}})}{Y_{ij}} \tag{8}$$

and the collective household model that yields (8) is

$$U_{ij} = a_{ij} + \gamma_{ij} \ln c_i, \qquad V_{ij} = b_{ij} + \gamma_{ij} \ln c_j, \qquad c_i + c_j = Y_{ij} \tag{9}$$

Here, a_{ij} and b_{ij} represent the marital gains for men and women, respectively, which can include both public goods and non-economic components of marriage. The parameter γ_{ij} determines the curvature of consumption in the utility function and, since utility transfers are made via private consumption, it also affects the curvature of the bargaining frontier. An interesting property of this model is that γ_{ij} reflects the degree of transferability: as γ_{ij} approaches $+\infty$, utility becomes perfectly transferable, whereas as γ_{ij} approaches zero, the model approximates a non-transferable utility (NTU) framework. We assume $\gamma_{ij} = \gamma + \epsilon_{ij}$, where ϵ_{ij} has a mean of zero and finite variance conditional on γ, such that as γ approaches zero or $+\infty$, the model transitions to the NTU and TU frameworks, respectively, for all types ij.

For singles, the utility function does not include the marital gain terms a_{ij} and b_{ij}. Instead, it is a logarithmic function of their consumption, which equals their own income. Instead of assuming separate scaling parameters γ_{i0} and γ_{0j} for single individuals, we assume that when deciding to match with a partner of a specific type, individuals use the same degree of transferability for their singlehood utilities as they do for their consumption when matched with a potential mate. In other words, the singlehood utilities that a man with education i and a woman with education j consider when deciding whether to match with each other are

$$U_{i0} = \gamma_{ij} \ln Y_{i0}, \qquad V_{0j} = \gamma_{ij} \ln Y_{0j} \tag{10}$$

In this setting, from (7), the Pareto weight of ETU model becomes

$$\lambda_{ij} = \frac{\exp(\frac{U_{ij}-a_{ij}}{\gamma_{ij}})}{\exp(\frac{U_{ij}-a_{ij}}{\gamma_{ij}}) + \exp(\frac{V_{ij}-b_{ij}}{\gamma_{ij}})} = \frac{1}{1 + \frac{Y_{0j}}{Y_{i0}} \left(\frac{N_{i0} \exp a_{ij}}{N_{0j} \exp b_{ij}} \right)^{\frac{1}{\gamma_{ij}}}} \tag{11}$$

According to Browning et al. (2014), Pareto weight is a *distribution factor* in the collective model, characterized by elements beyond preferences and budget constraints. The ETU framework allows us to endogenize this important parameter into the model. Specifically, in the marriage market equilibrium, the relative power of a man with education i when matched with a woman with education j is determined by three factors:

- The income ratio if single (Y_{i0}/Y_{0j}), which reflects the reservation utilities of singlehood and serves as a bargaining factor.
- The inverse of the population ratio of singles in their respective types (N_{0j}/N_{i0}), which indicates the availability of potential mates of the same type in the marriage market.
- The difference between marital gains ($b_{ij} - a_{ij}$), where the partner with lower non-monetary gains from marriage is compensated by receiving a greater Pareto weight in equilibrium.

In addition, we can compute the marriage surplus in the above model as

$$S_{ij} = \underbrace{\frac{1}{2}(a_{ij} + b_{ij})}_{\substack{\text{non-monetary} \\ \text{component}}} + \underbrace{\gamma_{ij} \ln \left(Y_{ij} \sqrt{\frac{\lambda_{ij}(1-\lambda_{ij})}{Y_{i0} Y_{0j}}} \right)}_{\text{monetary component}} \tag{12}$$

Therefore, the surplus increases with marital gains a_{ij} and b_{ij} as well as household income Y_{ij}, while it decreases with income if remaining single Y_{i0} and Y_{0j}. Additionally, the surplus is maximized when an even sharing rule is applied within households. Another implication of equation (13) is the decomposition of the marriage surplus into two components: a non-monetary component, which is independent of income, and a monetary component, which is determined by the return to education. Later, we will use this decomposition in our counterfactual exercises.

5.5 Parameter Identification

The above model has three parameters to identify for each couple type: a_{ij}, b_{ij}, and γ_{ij}. To identify martial gains, note that ETU model leads to this matching function for each couple type

$$N_{ij} = \left(\frac{Y_{i0}}{Y_{ij}} \left(N_{i0} e^{a_{ij}} \right)^{\frac{-1}{\gamma_{ij}}} + \frac{Y_{0j}}{Y_{ij}} \left(N_{0j} e^{b_{ij}} \right)^{\frac{-1}{\gamma_{ij}}} \right)^{-\gamma_{ij}} \tag{13}$$

and by combining (11) and (15), we obtain

$$a_{ij} = \gamma_{ij} \ln \frac{Y_{i0}}{\lambda_{ij} Y_{ij}} + \ln \frac{N_{ij}}{N_{i0}}, \qquad b_{ij} = \gamma_{ij} \ln \frac{Y_{0j}}{(1-\lambda_{ij}) Y_{ij}} + \ln \frac{N_{ij}}{N_{0j}} \tag{14}$$

Thus, upon having information on sharing rule λ_{ij}, marital gain parameters can be readily identified from (16), given the level of γ_{ij}. Still, we can identify the below lower-bounds for the marital gains

$$a_{ij} \geq \gamma_{ij} \ln \frac{Y_{i0}}{Y_{ij}} + \ln \frac{N_{ij}}{N_{i0}}, \qquad b_{ij} \geq \gamma_{ij} \ln \frac{Y_{0j}}{Y_{ij}} + \ln \frac{N_{ij}}{N_{0j}} \tag{15}$$

Theorem 5 of GKW shows that point-identification of the parameters a_{ij} and b_{ij} requires information on transfers between couples. Without this information, only set-identification of these parameters is possible for a given level of γ_{ij}. Since the CPS data does not provide information on these transfers, we need additional assumptions to identify these parameters.

From (11), for any level of γ_{ij}, we have

$$\lambda_{ij} \in \begin{cases} (\lambda^*_{ij}, 1] & \text{if } N_{i0} \exp(a_{ij}) < N_{0j} \exp(b_{ij}) \\ \lambda^*_{ij} & \text{if } N_{i0} \exp(a_{ij}) = N_{0j} \exp(b_{ij}) \\ [0, \lambda^*_{ij}) & \text{if } N_{i0} \exp(a_{ij}) > N_{0j} \exp(b_{ij}) \end{cases} \qquad \text{where } \lambda^*_{ij} = \frac{Y_{i0}}{Y_{i0} + Y_{0j}} \tag{16}$$

Under the non-transferable utility (NTU) case, where $\gamma_{ij} \to 0$, the Pareto weight can be 0, λ^*_{ij}, or 1, depending on the comparison between $N_{i0} \exp(a_{ij})$ and $N_{0j} \exp(b_{ij})$. In the TU case, where $\gamma_{ij} \to +\infty$, λ^*_{ij} is the only possible Pareto weight, regardless of the direction of the inequality in the condition. Therefore, a reasonable choice for the Pareto weight, regardless of γ_{ij}, is λ^*_{ij}. In the collective model described above, one can show that $c_i = \lambda_{ij} Y_{ij}$ and $c_j = (1-\lambda_{ij}) Y_{ij}$. Thus, the Pareto weight determines the private consumption sharing rule and does not affect the non-transferable component of utilities. When $\lambda_{ij} = \lambda^*_{ij}$, couples allocate their household income based on their potential income if they remained single. In equilibrium, the population ratio of singles will reflect the ratio of the non-monetary gains that are not transferable. Under this scenario, we also obtain

$$S_{ij} = \frac{1}{2}(a_{ij} + b_{ij}) + \gamma_{ij} \ln \frac{Y_{ij}}{Y_{i0} + Y_{0j}} \tag{17}$$

which suggests that the monetary part of the surplus is equal to the log of the ratio of couple income to the sum of single's income.

An alternative to using λ^*_{ij} as the sharing rule is to assume that λ_{ij} is a monotone function of γ_{ij}, such that λ_{ij} approaches λ^*_{ij} as $\gamma_{ij} \to +\infty$, and becomes either 0 or 1 as $\gamma_{ij} \to 0$. A reasonable assumption is that for couple types where men and women have more education, λ_{ij} is in $(\lambda^*_{ij}, 1]$ and $[0, \lambda^*_{ij})$, respectively. When both partners have the same education level, λ_{ij} would lie between λ^*_{ij} and 0.5. A simple functional form that captures this property is

$$\lambda_{ij} = \frac{\mathbb{1}(i > j)K + \mathbb{1}(i = j)K/2 + \gamma_{ij}\lambda^*_{ij}}{K + \gamma_{ij}} \tag{18}$$

Here, $K \geq 0$ is a scalar that we can try different levels of it for checking the robustness of the results.

According to Theorem 5 of GKW, identification of γ_{ij} requires information on transfer across multiple markets. Due to data limitation on transfer, we determine γ_{ij} by leveraging the homoskedasticity of random terms in utilities. We choose γ_{ij} such that the monetary component of the household utility also becomes homoskedastic. According to Proposition 2 and Assumption 2, the stochastic part of the surplus (19) is $\frac{1}{2}(\alpha^j_m - \alpha^0_m + \beta^i_f - \beta^0_f)$ which is the average of two standard logistic random variables. Assuming log-normal distribution for the income each couple type, we choose γ_{ij} such that the variance of the monetary term of the surplus in (19) equals $\pi^2/3$ which is the variance of a standard logistic random variables. This yields

$$\gamma_{ij} = \frac{\pi}{\sqrt{3\mathrm{Var}(\ln Y_{ij} - \ln Y_{i0} - \ln Y_{0j})}} \tag{19}$$

5.6 Finding Equilibrium Matching

To characterize the equilibrium matching N_{ij} in the marriage market, we need estimates of the non-monetary gains (a_{ij} and b_{ij}), the transferability parameter (γ_{ij}), and data on the average income of couples (Y_{ij}) and singles (Y_{i0} and Y_{0j}), as well as the marginal population vectors ($N_{i\oplus}$ and $N_{\oplus j}$). While Algorithm 2 of GKW provides a method for finding equilibrium in a general ITU framework, we employ a more efficient algorithm that avoids solving a system of non-linear equations.

Given non-monetary and monetary gains, first note that N_{i0} and N_{0j} are the solutions to the below equations

$$N_{i0} + \sum_{j=1}^{J} \exp(S_{ij})\sqrt{N_{i0}N_{0j}} = N_{i\oplus}, \qquad N_{0j} + \sum_{i=1}^{I} \exp(S_{ij})\sqrt{N_{i0}N_{0j}} = N_{\oplus j} \tag{20}$$

which yield

$$N_{i0}(S_{i\bullet}, N_{0\bullet}, N_{i\oplus}) = \frac{1}{4}\Big(\sqrt{\big(\sum_{j=1}^{J}\exp(S_{ij})\sqrt{N_{0j}}\big)^2 + 4N_{i\oplus}} - \sum_{j=1}^{J}\exp(S_{ij})\sqrt{N_{0j}}\Big)^2 \tag{21}$$

$$N_{0j}(S_{\bullet j}, N_{\bullet 0}, N_{\oplus j}) = \frac{1}{4}\Big(\sqrt{\big(\sum_{i=1}^{I}\exp(S_{ij})\sqrt{N_{i0}}\big)^2 + 4N_{\oplus j}} - \sum_{i=1}^{I}\exp(S_{ij})\sqrt{N_{i0}}\Big)^2 \tag{22}$$

In addition, the marriage surplus can be written as a function of marital gains, income matrices, and singles' populations as

$$S_{ij}(N_{i0}, N_{0j}) = \gamma_{ij}\ln Y_{ij} - \gamma_{ij}\ln\Big(Y_{i0}\big(\frac{N_{0j}}{N_{i0}}\big)^{\frac{1}{2\gamma_{ij}}}\exp(-\frac{a_{ij}}{\gamma_{ij}}) + Y_{0j}\big(\frac{N_{i0}}{N_{0j}}\big)^{\frac{1}{2\gamma_{ij}}}\exp(-\frac{b_{ij}}{\gamma_{ij}})\Big) \tag{23}$$

Using these functions, for initial values of singles' population, we follow the below iterative procedure until it converges to the equilibrium:

1. $S_{ij}^{(k)} = S_{ij}(N_{i0}^{(k)}, N_{0j}^{(k)})$
2. $N_{i0}^{(k+1)} = N_{i0}(S_{i\bullet}^{(k)}, N_{0\bullet}^{(k)})$
3. $N_{0j}^{(k+1)} = N_{0j}(S_{\bullet j}^{(k)}, N_{\bullet 0}^{(k+1)})$

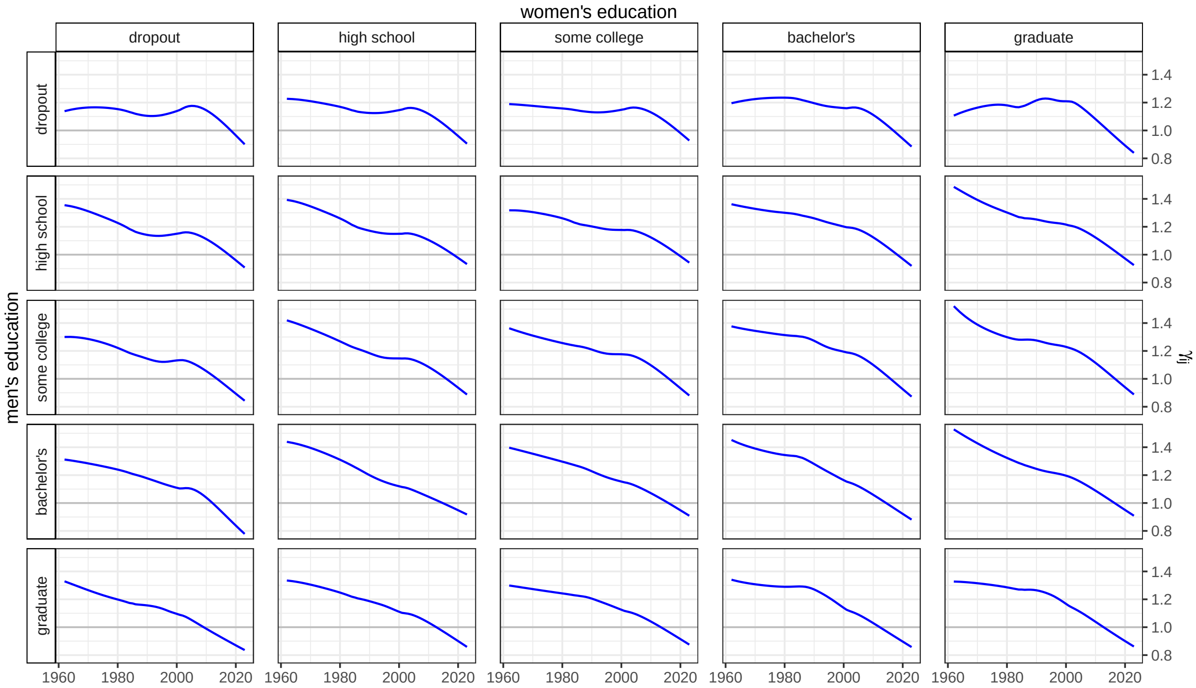
Then, we can simply find $N_{ij} = \exp(S_{ij})\sqrt{N_{i0}N_{0j}}$.

6 Estimation

In this section, we present the estimated parameters and the procedure to build counterfactual experiments.

As illustrated in the appendix Figure 14, couple types with low sample sizes exhibit fluctuations in average income and its variance over time. For this reason, to estimate γ_{ij} from (21), we use the smoothed version of $\mathrm{Var}(\ln Y_{ij})$ through non-parametric LOESS regression. The top plot of Figure 7, shows our estimation for γ_{ij} varies between 0.8 and 1.5 across different groups. Moreover, we observe a decreasing trend over time in almost all couple types, suggesting higher income variance within each group in recent years.

We illustrate the trend of λ_{ij}^* in bottom plot of Figure 7. The numbers indicate that the sharing rule favors men in the lower-left region and favors women in the upper-right region. Thus, higher education is associated with greater bargaining power within the family. Along the diagonal, we observe a slightly higher income share for men. In the following estimations, we derive λ_{ij} using equation 20) with $K = 1$.



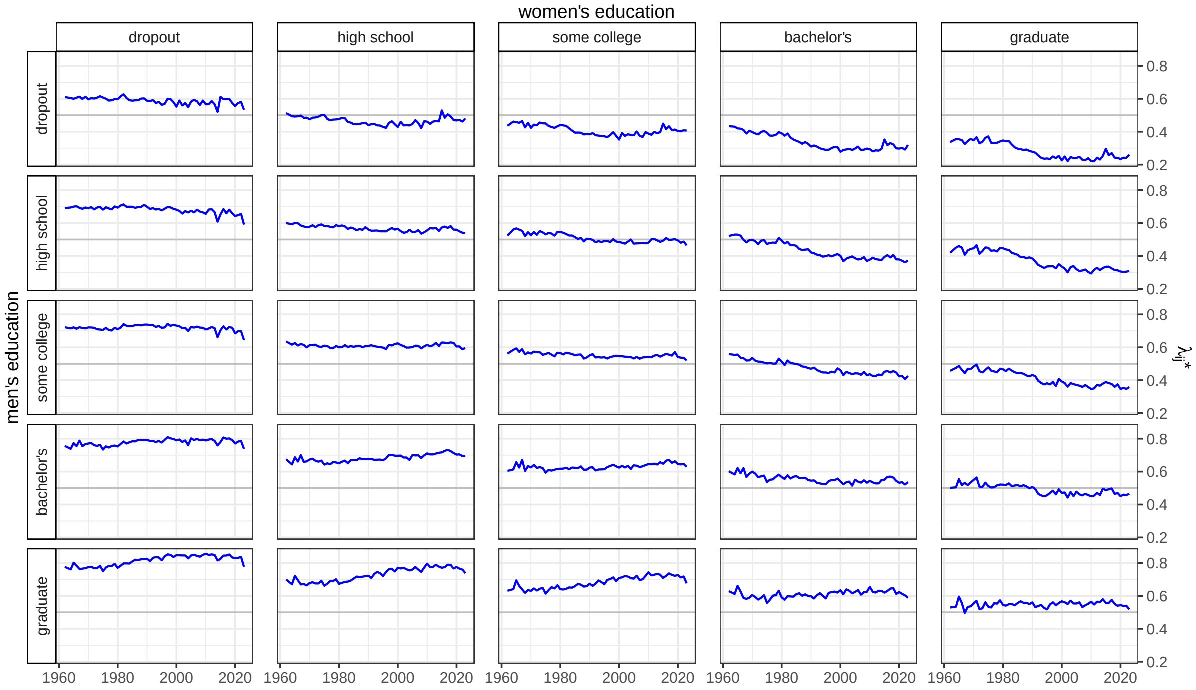


Figure 7: The estimated levels of γ_{ij} and λ^*_{ij}.

6.1 Exogenous vs. Endogenous Non-monetary Gains

Given the levels of λ_{ij} and γ_{ij}, and with observations of population and income at a specific point in time, one can estimate a_{ij} and b_{ij} using equation (16). However, the non-monetary gains can be decomposed into two components based on their dependence on population marginals. The exogenous component, which relates to factors such as affinity and public goods, is generally independent of changes in $N_{i\oplus}$ and $N_{\oplus j}$. In contrast, the non-economic gains also include a component that depends on competition in the marriage market. This component is influenced by the number of similar individuals and potential mates within a specific marriage type, which are determined by secular trends in the marginal populations $N_{i\oplus}$ and $N_{\oplus j}$. These trends are driven by changes in the demand and supply of education, such as increased labor demand for educated individuals and pre-matching investments in education. Since our goal is to conduct counterfactual experiments while fixing exogenous factors, we need to separate the first component from the component influenced by the educational distribution vectors.

To decompose these two components, we consider the hypothetical case of random matching, where the marriage surplus is zero for all types, and individuals are indifferent between marriage and singlehood. In this scenario, the population in each couple type is the geometric mean of their respective single populations: $\bar{N}_{ij} = \sqrt{\bar{N}_{i0}\bar{N}_{0j}}$, where $\bar{N}_{i0}$ and $\bar{N}_{0j}$ are the single populations under random matching. Given the educational distribution vectors $N_{i\oplus}$ and $N_{\oplus j}$, the single population vectors under random matching, $\bar{N}_{i0}(N_{\cdot\oplus}, N_{\oplus\cdot})$ and $\bar{N}_{0j}(N_{\cdot\oplus}, N_{\oplus\cdot})$, are determined by solving the below system of equations using the iterative algorithm described in section 5.6

$$\bar{N}_{i0} + \sum_{j=1}^{J} \sqrt{\bar{N}_{i0}\bar{N}_{0j}} = N_{i\oplus} \qquad \bar{N}_{0j} + \sum_{i=1}^{I} \sqrt{\bar{N}_{i0}\bar{N}_{0j}} = N_{\oplus j} \tag{24}$$

In this regard, we define the below modified non-economic gains that are adjusted for the population distribution changes

$$\tilde{a}_{ij} = a_{ij} + \frac{1}{2}\ln\frac{\bar{N}_{i0}}{\bar{N}_{0j}} = \gamma_{ij}\ln\frac{Y_{i0}}{\lambda_{ij}Y_{ij}} + \ln\frac{N_{ij}\bar{N}_{i0}}{\bar{N}_{ij}N_{i0}} \tag{25}$$

$$\tilde{b}_{ij} = b_{ij} + \frac{1}{2}\ln\frac{\bar{N}_{0j}}{\bar{N}_{i0}} = \gamma_{ij}\ln\frac{Y_{0j}}{(1-\lambda_{ij})Y_{ij}} + \ln\frac{N_{ij}\bar{N}_{0j}}{\bar{N}_{ij}N_{0j}} \tag{26}$$

The components of a_{ij} and b_{ij} in (16) that are endogenous to population supplies are $\ln N_{ij} - \ln N_{i0}$ and $\ln N_{ij} - \ln N_{0j}$, respectively. By modifying these terms to $\ln\frac{N_{ij}}{\bar{N}_{ij}} - \ln\frac{N_{i0}}{\bar{N}_{i0}}$ and $\ln\frac{N_{ij}}{\bar{N}_{ij}} - \ln\frac{N_{0j}}{\bar{N}_{0j}}$, we adjust for the effects of population marginals $(N_{\cdot\oplus}, N_{\oplus\cdot})$ by computing the deviations from the random matching case, rather than using their absolute values, which are influenced by population marginals. Note that the sum of the adjusted and unadjusted non-monetary terms are equal: $a_{ij} + b_{ij} = \tilde{a}_{ij} + \tilde{b}_{ij}$. Thus, the non-monetary component of the surplus in (13) can also be measured using the adjusted non-monetary gains.

Figure 8, illustrates the estimation of the adjusted non-monetary parameters $\tilde{a}_{ij}$ and $\tilde{b}_{ij}$ using (16) and compares them with the average surplus $S_{ij} = \ln N_{ij} - \frac{1}{2}\ln(N_{i0}N_{0j})$. A similar figure for the unadjusted parameters is available in the Appendix. We observe that the estimated values of non-monetary gains generally align with the average surplus. The diagonal elements, where couples assortatively match by education, have high surplus levels. Conversely, the values are negative for anti-diagonal elements, indicating that matches between partners with different education levels tend to be less desirable on average. Additionally, in most cases, the lower-educated partner receives more non-monetary gain compared to the higher-educated partner, reflecting the trade-off that higher-educated partners typically obtain a larger share of economic gains.

6.2 Building Counterfactual

In this section, we outline the procedure for conducting counterfactual experiments to analyze marriage market outcomes and income inequality in the US over the period 1962-2003. Our first set of exercises examines the contribution of three exogenous factors on the outcomes: population marginals, adjusted non-monetary gains, and the average income matrix by education and marital status. For these counterfactuals, we select three base years to establish benchmark levels for these components:

- t_N: time of measuring the marginal distribution of population by education ($N_{i\oplus}$ and $N_{\oplus j}$)
- t_X: time of measuring the adjusted non-monetary benefit matrices ($\tilde{a}_{ij}$ and $\tilde{b}_{ij}$)
- t_Y: time of measuring the monetary benefits measured by income matrix Y_{ij}

For population and income, we only change the distribution of the variables by type as the benchmark levels and the sum of population and total income is always at their current levels. After choosing these benchmark level, we first estimate $\bar{N}_{i0}^{t_N}$ and $\bar{N}_{0j}^{t_N}$ from (27) and using them we obtain

$$a_{ij}(t_N, t_X) = \tilde{a}_{ij}^{t_X} + \frac{1}{2}\ln\frac{\bar{N}_{0j}^{t_N}}{\bar{N}_{i0}^{t_N}} \qquad b_{ij}(t_N, t_X) = \tilde{b}_{ij}^{t_X} + \frac{1}{2}\ln\frac{\bar{N}_{i0}^{t_N}}{\bar{N}_{0j}^{t_N}}$$

Then, using income matrix of t_Y, we specify the surplus as a function of single population as in (26), and follow the algorithm of section 5.6 to estimate the counterfactual population matrix, marriage market outcomes and the Gini index.

6.3 Counterfactuals with constant marriage market outcomes

The second set of counterfactual exercises involves fixing the marriage market outcomes to benchmark levels. This exercise can be on fixing either AM, marriage rate, or both at a benchmark time and then computing the counterfactual population and income matrices at a subsequent time. For the population matrix, after specifying the desired levels of the matrix ρ and vectors μ and ω, Proposition 1 can be used to determine the values of N_{ij}, N_{i0}, N_{0j}. However, determining the income matrix is more complex. To

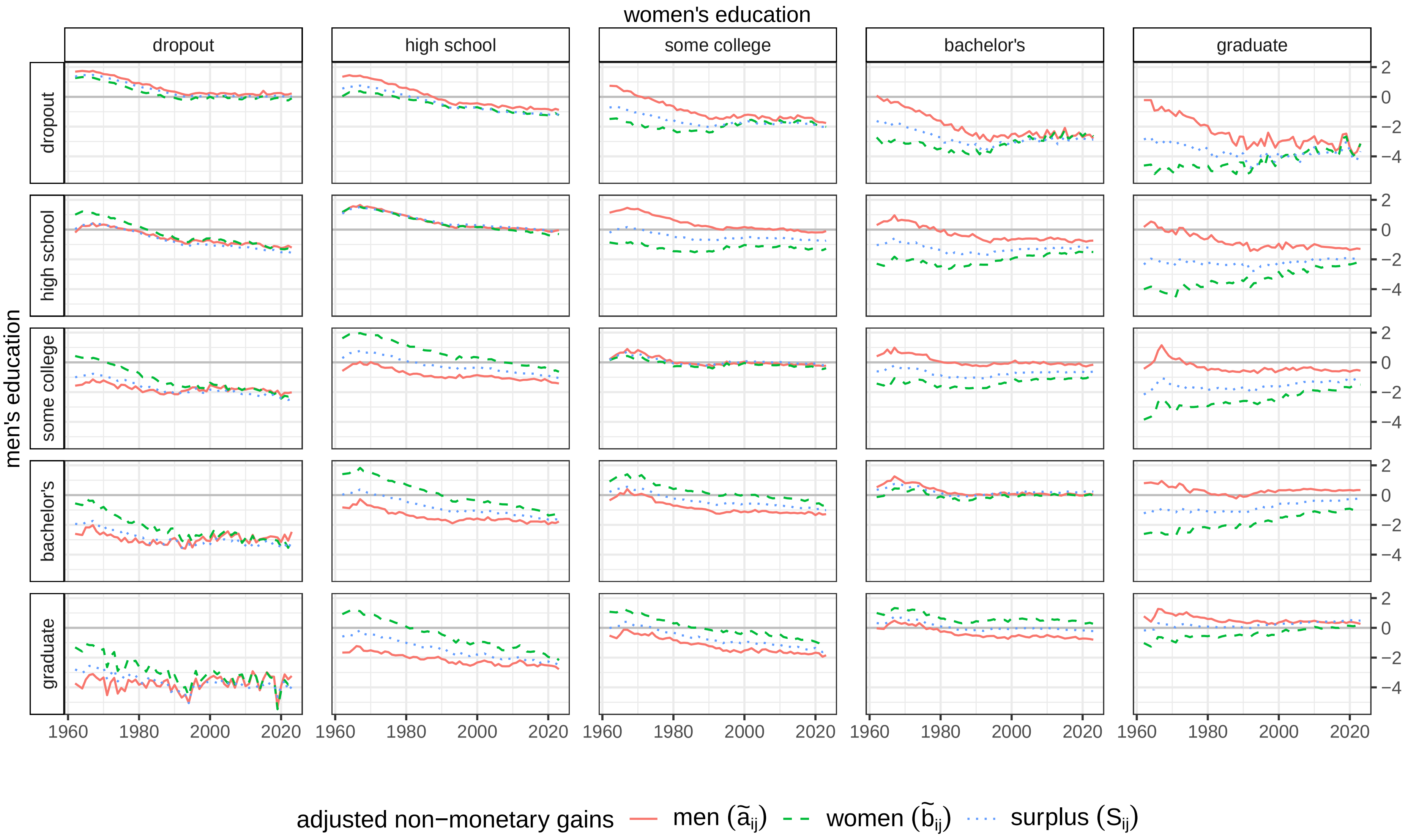


Figure 8: The estimated levels of non-monetary gains a_{ij} and b_{ij}, and total surplus S_{ij}.

characterize it, we first define

$$C_{i\oplus} := N_{i0}Y_{i0} + \sum_{j=1}^{J} N_{ij}\lambda_{ij}Y_{ij}, \qquad C_{\oplus j} := N_{0j}Y_{0j} + \sum_{i=1}^{I} N_{ij}(1-\lambda_{ij})Y_{ij} \tag{27}$$

Here, $C_{i\oplus}$ and $C_{\oplus j}$ are total consumption by each gender and education level in the whole economy that we assume that in the counterfacutal exercises it is at its current level. Indeed, $C_{i\oplus}/N_{i\oplus}$ and $C_{\oplus j}/N_{\oplus j}$ are return to education levels i and j for men and women, respectively. In addition, we define two variables for the share of income of married people by education

$$\phi_i := 1 - \frac{N_{i0}Y_{i0}}{C_{i\oplus}}, \qquad \varphi_j := 1 - \frac{N_{0j}Y_{0j}}{C_{\oplus j}} \tag{28}$$

To build a counterfactual experiment, in addition to AM and marriage rate, we need to choose the desired level of married income shares (ϕ_i) and (φ_j), too. These parameters enable us to find Y_{i0} and Y_{0j} from (31) and also λ_{ij} from (20).

Still, we need to characterize the couple's income Y_{ij} in a counterfactual scenario. Let, $\Delta(.)$ be the demean operator defined on a two-way variable X_{ij} as

$$\Delta(X_{ij}) = X_{ij} - \frac{1}{J}X_{i+} - \frac{1}{I}X_{+j} + \frac{1}{IJ}X_{++}$$

Note that, because the row and column sums of demeaned components are zero, the demeaned elements ΔX_{ij} cannot determine the elements of X_{ij} without knowing the row and column means. According to Proposition 4, AM elements are the demeaned values of the surplus $\rho_{ij} = \Delta(S_{ij})$. Therefore, from (13), we have

$$\rho_{ij} = \frac{1}{2}\Delta(a_{ij}+b_{ij}) + \frac{1}{2}\Delta\Big(\gamma_{ij}\ln\frac{\lambda_{ij}(1-\lambda_{ij})}{Y_{i0}Y_{0j}}\Big) + \Delta\Big(\gamma_{ij}\ln Y_{ij}\Big) \tag{29}$$

Therefore, in a scenario that AM is constant, we need to decide about the time to measure a_{ij}, b_{ij}, and γ_{ij}, too. Then, from (32), we can obtain the odd ratios of a matrix with elements $Y_{ij}^{\gamma_{ij}}$ as

$$\delta_{ij} := \exp\big(\Delta(\gamma_{ij}\ln Y_{ij})\big) = \frac{Y_{ij}^{\gamma_{ij}}\prod_{k=1}^{I}\prod_{l=1}^{J}Y_{kl}^{\gamma_{kl}}}{\prod_{l=1}^{J}Y_{il}^{\gamma_{il}}\prod_{k=1}^{I}Y_{kj}^{\gamma_{kj}}}$$

In this regard, we have $Y_{ij}^{\gamma_{ij}} = d_{1i}\delta_{ij}d_{2j}$ (see the proof of Proposition 1), where the vectors of d_{1i} and d_{2j} must be chosen to satisfy the below row and column sums conditions

$$\sum_{j=1}^{J} N_{ij}\lambda_{ij}Y_{ij} = \phi_i C_{i\oplus}, \qquad \sum_{i=1}^{I} N_{ij}(1-\lambda_{ij})Y_{ij} = \varphi_i C_{\oplus j} \tag{30}$$

Hence, we determine d_{1i} and d_{2j} by solving the below least square estimator

$$\min_{d_1,d_2} \sum_{i=1}^{I}\Big(\sum_{j=1}^{J} N_{ij}\lambda_{ij}(d_{1i}\delta_{ij}d_{2j})^{\frac{1}{\gamma_{ij}}}/\phi_i C_{i\oplus} - 1\Big)^2 + \sum_{j=1}^{J}\Big(\sum_{i=1}^{I} N_{ij}(1-\lambda_{ij})(d_{1i}\delta_{ij}d_{2j})^{\frac{1}{\gamma_{ij}}}/\varphi_j C_{\oplus j} - 1\Big)^2 \tag{31}$$

In summary, to perform a counterfactual exercise with given values of marriage market outcomes, we proceed as follows:

1. Choose the timing of these parameters in counterfactual scenario: ρ, μ, ω, ϕ, φ, a, b, and γ.
2. From Proposition 1, estimate population matrix using $N_{\cdot\oplus}$, $N_{\oplus\cdot}$, μ, ω, and ρ.
3. Using $C_{\cdot\oplus}$, $C_{\oplus\cdot}$,ϕ, φ, and $N_{\cdot 0}, N_{0\cdot}$ from step 1, estimate vectors of $Y_{\cdot 0}$ and $Y_{0\cdot}$.
4. Using $Y_{\cdot 0}$ and $Y_{0\cdot}$, find λ^* and from (20) estimate λ.
5. Find δ from (32), given ρ, a, b, γ, and the estimation of step 4.
6. Estimate d_1 and d_2 from (34)
7. Compute $Y_{ij} = (d_{1i}\delta_{ij}d_{2j})^{\frac{1}{\gamma_{ij}}}$.

7 Findings

In this section, we present the results of our counterfactual exercises using CPS data for 1962-2023. The first set of scenarios involves fixing one of t_N, t_X, and t_Y to their 1962 levels, and also exploring the scenario where all three are fixed simultaneously. We first describe the results regarding the marriage market outcomes, and then present the counterfactual findings regarding cross-sectional income inequality in the US over 1962-2023. Finally, we will analyze the trend of counterfactual inequality when AM is maintained at its 1962 level.

7.1 Counterfactual AM and Marriage Rates

Figure 9 illustrates the counterfactual trends of the average marriage rate under various scenarios for men and women. For detailed trends of marriage rates by education level, refer to Appendix Figure 16. The counterfactual scenarios where income or population patterns are fixed at their 1962 levels both indicate a lower marriage rate compared to the actual observed levels. In contrast, the scenario where non-monetary gains are held constant at 1962 levels suggests that the marriage rate would have been higher today if the pattern of non-pecuniary gains had remained unchanged. This aligns with the decreasing trend of surplus depicted in Figure 8. Additionally, these counterfactual trends do not exhibit the 1995 break (the

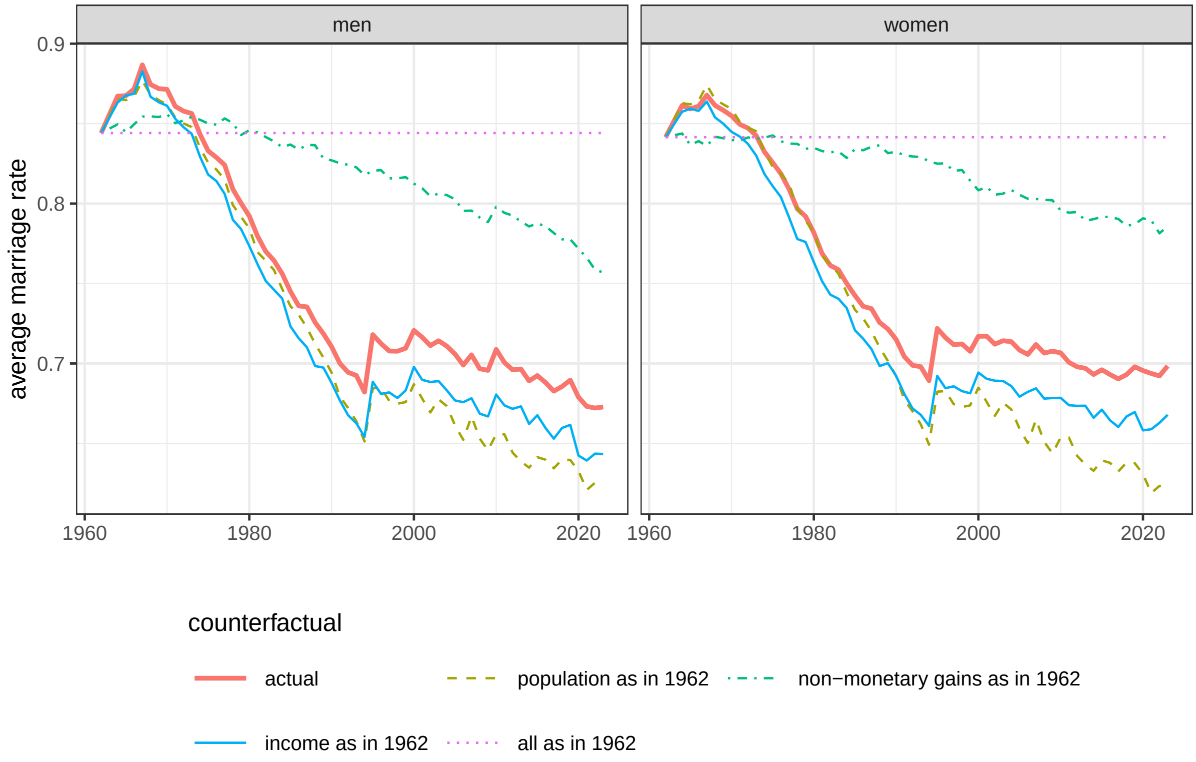


Figure 9: The counterfactual trends of average marriage rate in different scenarios.

onset of cohabitation data) because this break is accounted for by the estimated trends of non-monetary gains.

Figure 10 shows the trend of aggregate AM measure under different scenarios. We present only the aggregate measure using the weighted log odds ratio index and trends for other indices and individual log odds ratios can be found in the Appendix. The figure reveals significant changes in the AM trend across all scenarios. If the population distribution had remained unchanged since 1962, the aggregate AM would have experienced much smaller fluctuations compared to its actual trajectory. The impact of variations in non-monetary gains on AM is significant before 1990, while after this period, the trend remains relatively similar to the actual one. Lastly, if the return to education had remained at its 1962 level, the AM would have been higher before 1990 and lower thereafter compared to its current level.

7.2 The Counterfactual US Income Inequality

Figure 12 shows the estimated Gini coefficient under four counterfactual scenarios in the US: when either population, non-monetary gains, or income is the same as 1962, and when all three factors are as the levels of 1962. The top graph shows the estimated inequality for all household and the middle graphs show them for the sample of married households. In the bottom graph of Figure 12, we estimate the Gini coefficient at the individual level, by dividing the household income of the married couples of each type ij between the partners, according to their average sharing rule λ_{ij}. On average, inequality is highest

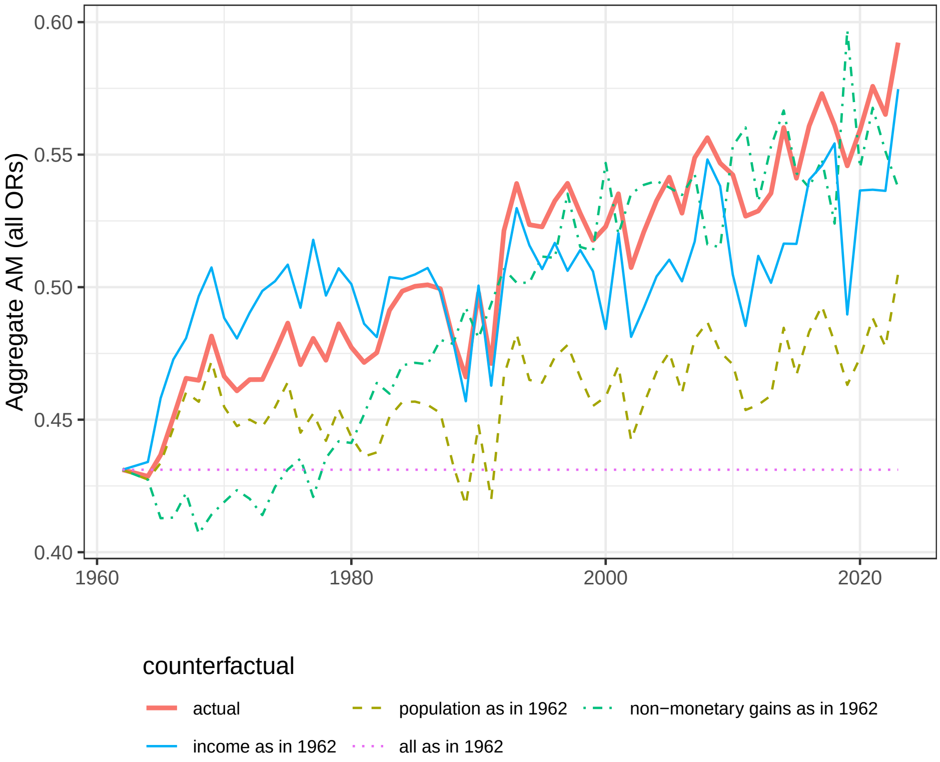


Figure 10: The counterfactual trends of marriage return in different scenarios.

in the sample of all individuals, particularly in the earlier years of the data. However, with the decline in marriage rates post-2000, the Gini index for all households converges to similar levels as that for all individuals. Throughout all years, married couples exhibit the lowest level of inequality.

We first examine the counterfactual scenario where the population distribution is fixed at its 1962 level. In this experiment, we find that inequality would have been higher in all years across all samples. This suggests that the secular trends in educational attainment have contributed to reducing income inequality. In other words, as the return to education increased, more individuals pursued additional schooling, which, in turn, reduced inequality relative to the initial distribution. The largest discrepancies between actual and counterfactual trends occur in the 1980s and 1990s, a period during which, as shown in Figure 1, there was a significant shift in educational attainment from low to high levels, particularly among men. By the end of the study period, the trends suggest that if the educational distribution had remained at its 1962 level, the Gini index in 2023 would have been 2 Gini points (4%) higher for both the sample of all households and individuals.

In the counterfactual scenario where t_X is fixed at 1962, we observe little changes in inequality for the samples of married couples and individuals, but a significant reduction in the Gini index for the sample of all households. Combining this with the counterfactual results for the marriage rate (Figure 9) suggests that the main driver behind the reduction in the Gini coefficient for all households is the higher marriage rate when non-monetary gains are set to their 1962 levels. When the marriage rate increases, and inequality is measured at the household level by including both single and married households, a higher rate results in a more homogeneous sample of households in terms of household size, thus leading to lower inequality. Since the other factors do not affect marriage rate as significantly as non-monetary gains do, their counterfactual trends show relatively little variation across different samples.

In the third scenario, where $t_Y = 1962$, we observe a significant decline in inequality, amounting to roughly 4 Gini points in the final years of the data. In 1962, the Gini coefficients were 0.378 for all households, 0.328 for married couples, and 0.404 for all individuals. According to this counterfactual analysis, fixing the monetary gains from marriage accounts for 41%, 43%, and 53% of the overall increase in income inequality between 1962 and 2023 for the samples of all households, married couples, and all individuals, respectively.

Finally, when all exogenous factors are set to their 1962 levels, we still observe a decline in income inequality. This decline is the cumulative effect of changes attributable to the three factors. Under this scenario, the observed changes in inequality compared to the actual levels are due to shifts in income distribution within each household type, rather than between types. Consistent with this finding, Appendix Table A.2, shows an overall increase in the Gini index within groups in recent years compared to 1962.

7.3 The US Income Inequality with Constant AM

Figure 12 displays the estimated Gini coefficient under three counterfactual scenarios for the AM and compares it with the actual trend. Since AM is defined conditional on marriage, the trends are illustrated for married couples. The three scenarios are based on the components of AM that are benchmarked to their 1962 levels. According to (32), AM can be decomposed into non-monetary and monetary components as follows:

$$\rho_{ij} = \underbrace{\Delta\big(\frac{a_{ij} + b_{ij}}{2}\big)}_{\substack{\text{non-monetary} \\ \text{component}}} + \underbrace{\Delta\Big(\gamma_{ij} \ln\big(Y_{ij}\sqrt{\frac{\lambda_{ij}(1-\lambda_{ij})}{Y_{i0}Y_{0j}}}\big)\Big)}_{\text{monetary component}}$$

The first case involves fixing both components of AM to their 1962 levels and finding equilibrium according to Section 6.3. In this scenario, we observe a significant decline in inequality, approximately 3 Gini points, for married couples. In 1962, the Gini coefficient was 0.328 and based on this counterfactual, controlling for both monetary and non-monetary components that characterize AM accounts for 33 percent of the overall increase in income inequality between 1962 and 2023 for married couples.

In the counterfactual scenario where only the non-monetary component is fixed, we observe little increase in inequality, with the Gini coefficient rising by just 0.003 points for married couples in 2023. In contrast, when only the monetary component of AM is fixed, the trend closely resembles that observed when both components are set to their 1962 levels. This suggests that the primary driver of changes in income inequality due to AM is the monetary channel rather than the non-monetary channel.

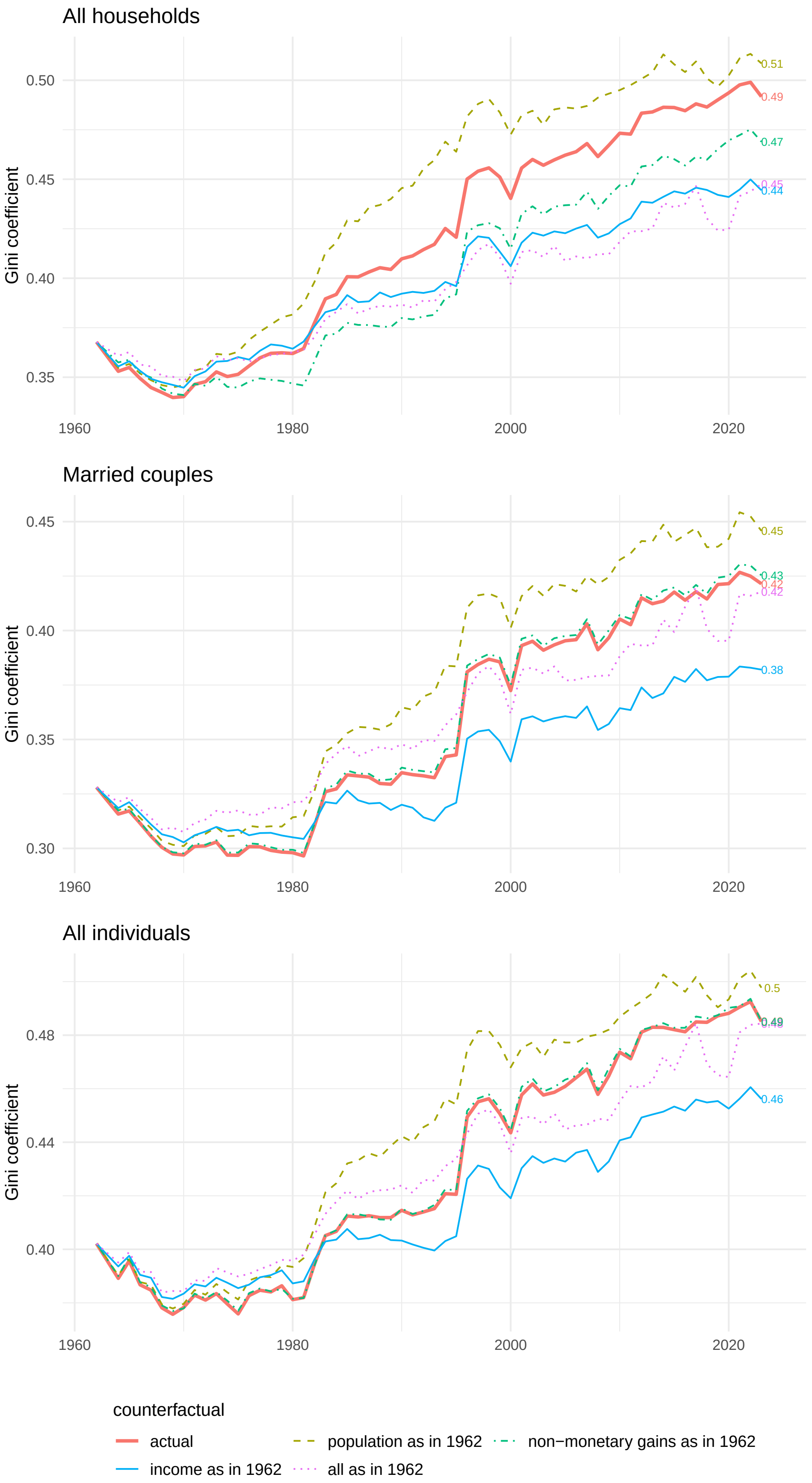


Figure 11: Counter-factual inequality in different scenarios. The top and middle graphs are at household level and in the bottom graph married couple of each type are counted as two individual with income according to the sharing rule λ_{ij}.

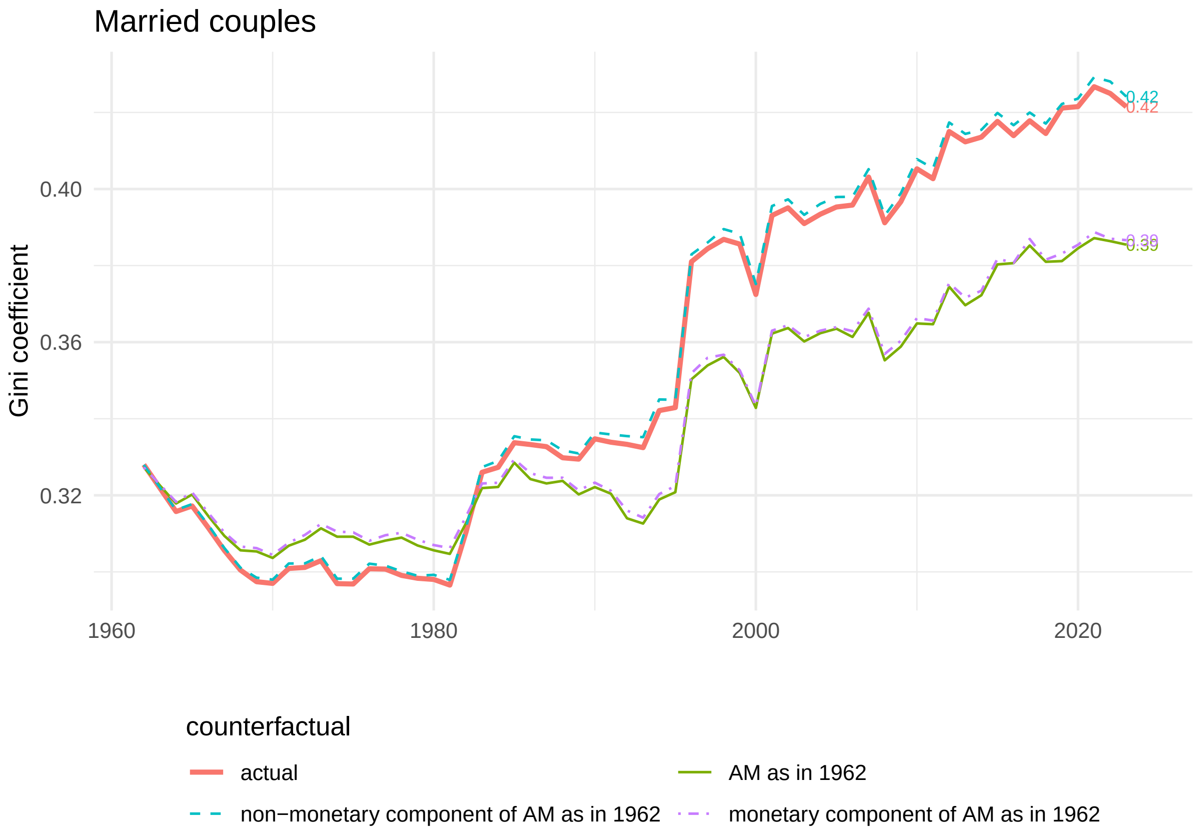


Figure 12: Counter-factual analysis with fixed AM.

8 Conclusion

This paper reexamines the relationship between marriage market outcomes and cross-sectional income inequality, focusing on the role of return to education in shaping both. To connect assortative mating (AM) with average income, we highlight the influence of secular trends in population and income as confounding factors. Furthermore, we develop a frictionless matching model with imperfectly transferable utility to link the return to education with the monetary gains from marriage which is an essential factor in determining equilibrium matching and its effect on inequality.

Our findings diverge from the existing literature, which suggests a minimal impact of AM growth on the rise in income inequality. Using CPS data for the US from 1962 to 2023, we find that changes in the average income by marriage type explain roughly 40% of the increase in the Gini coefficient. Moreover, if the monetary component driving AM were fixed at its 1962 level, this would account for approximately 33% of the rise in the Gini coefficient among married couples over the same period.

References

Altham, P. M. E. (1970): "The Measurement of Association of Rows and Columns for an $r \times s$ Contingency Table," *Journal of the Royal Statistical Society: Series B (Methodological)*, 32, 63–73.

Autor, D. H. (2014): "Skills, education, and the rise of earnings inequality among the "other 99 percent"," *Science*, 344, 843–851.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81, 813–846.

——— (1974): "A Theory of Marriage: Part II," *Journal of Political Economy*, 82, S11–S26.

Breen, R. and L. Salazar (2011): "Educational assortative mating and earnings inequality in the United States," *American Journal of Sociology*, 117, 808–843.

Browning, M., P.-A. Chiappori, and Y. Weiss (2014): *Economics of the Family*, Cambridge University Press.

Chiappori, P.-A. (1992): "Collective labor supply and welfare," *Journal of political Economy*, 100, 437–467.

Chiappori, P.-A., M. Costa-Dias, S. Crossman, and C. Meghir (2020a): "Changes in assortative matching and inequality in income: Evidence for the UK," *Fiscal Studies*, 41, 39–63.

Chiappori, P.-A., M. Costa Dias, and C. Meghir (2020b): "Changes in Assortative Matching: Theory and Evidence for the US," Tech. Rep. w26932, National Bureau of Economic Research, Cambridge, MA.

——— (2021): "The measuring of assortativeness in marriage: a comment," *Working Paper*.

Chiappori, P.-A. and E. Gugl (2020): "Transferable utility and demand functions," *Theoretical Economics*, 15, 1307–1333.

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

Choo, E. and A. Siow (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114, 175–201.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

Dupuy, A. and S. Weber (2022): "Marriage Market Counterfactuals Using Matching Models," *Economica*, 89, 29–43.

Eika, L., M. Mogstad, and B. Zafar (2019): "Educational Assortative Mating and Household Income Inequality," *Journal of Political Economy*, 127, 2795–2835.

Fernandez, R., N. Guner, and J. Knowles (2005): "Love and money: A theoretical and empirical analysis of household sorting and inequality," *The Quarterly Journal of Economics*, 120, 273–344.

Fernández, R. and R. Rogerson (2001): "Sorting and long-run inequality," *The Quarterly Journal of Economics*, 116, 1305–1341.

Fortin, N., T. Lemieux, and S. Firpo (2011): "Decomposition methods in economics," in *Handbook of labor economics*, Elsevier, vol. 4, 1–102.

Galichon, A., S. D. Kominers, and S. Weber (2019): "Costly concessions: An empirical framework for matching with imperfectly transferable utility," *Journal of Political Economy*, 127, 2875–2925.

Goldin, C. and L. F. Katz (2009): *The race between education and technology*, harvard university press.

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

——— (2015): "Corrigendum to Marry Your Like: Assortative Mating and Income Inequality," *https://www.jeremygreenwood.net/papers/ggksPandPcorrigendum.pdf*.

Hoseini, M. (2023): "Marriage and employment returns to female education," *Working paper.*

Hryshko, D., C. Juhn, and K. McCue (2017): "Trends in earnings inequality and earnings instability among US couples: How important is assortative matching?" *Labour Economics*, 48, 168–182.

Kateri, M. (2014): "Contingency table analysis," *Methods and implementation using R (First edition). Aachen, Germany: Editorial Advisory Booard.*

Kremer, M. (1997): "How much does sorting increase inequality?" *The Quarterly Journal of Economics*, 112, 115–139.

Larrimore, J., R. V. Burkhauser, S. Feng, and L. Zayatz (2008): "Consistent cell means for topcoded incomes in the public use March CPS (1976–2007)," *Journal of Economic and Social Measurement*, 33, 89–128.

Sinkhorn, R. (1967): "Diagonal equivalence to matrices with prescribed row and column sums," *The American Mathematical Monthly*, 74, 402–405.

Stephan, F. F. (1942): "An iterative method of adjusting sample frequency tables when expected marginal totals are known," *The Annals of Mathematical Statistics*, 13, 166–178.

A Appendix

A.1 Proofs

Proposition 1. Sinkhorn (1967)'s theorem states that if A is an $I \times J$ matrix with positive elements, given two positive vectors R and C of size I and J, such that $\sum_1^I r_i = \sum_1^J c_j$, there exists a unique matrix B of the form $D_1 A D_2$ such that D_1 and D_2 are $I \times I$ and $J \times J$ diagonal matrices, and the row and column sums of B are the elements of R and C.

Note that, because $b_{ij} = d_{1i} a_{ij} d_{2j}$ in this theorem, the odds ratios in A and B are equal. Using this theorem, we can show that given the vectors $N_{i+} = \mu_i N_{i\oplus}, N_{+j} = \omega_j N_{\oplus j}$, we can determine $N_{ij},\ i,j > 0$ using a basic set of assortative matching terms. For instance, let $\rho^1_{ij} = \ln \frac{N_{11}}{N_{1j}} \frac{N_{ij}}{N_{i1}}$ be the nominal first set and consider A as an $I \times J$ matrix that its first row and column are ones and the remaining $(I-1) \times (J-1)$ submatrix contains $\exp(\rho^1_{ij}),\ i,j > 1$. Let R and C the vectors of N_{i+} and N_{+j}, respectively. Then, according to Sinkhorn's theorem, the unique matrix B will include $N_{ij},\ i,j > 0$. Because, any basic sets of odds ratios are convertible to another, this proposition applies to all such basic sets. □

Proposition 2. Since $\mathcal{B}_{ij}$ is a proper bargaining set, it has a distance-to-frontier function $D_{ij}(u, v)$. Using $D_{ij}(\cdot, \cdot)$, we can reformulate the stability conditions based on Assumption 1 as

- $\forall m \in i$, and $f \in j:\ D_{ij}(u_m - \alpha^j_m, v_f - \beta^i_f) \geq 0$ with equality when $\nu_{mf} = 1$
- $\forall m \in i : u_m \geq U_{i0} + \alpha^j_m$ with equality if $\sum_f \nu_{mf} = 0$ and $\forall f \in j : v_f \geq V_{0j} + \beta^i_f$ with equality if $\sum_m \nu_{mf} = 0$.

Consider $m, m' \in i$ and $f, f' \in j$ such that under stable matching m and m' respectively match with f and f'. From stability condition, we have

$$D_{ij}(u_m - \alpha^j_m, v_f - \beta^i_f) = 0 \qquad D_{ij}(u_{m'} - \alpha^j_{m'}, v_{f'} - \beta^i_{f'}) = 0$$
$$D_{ij}(u_{m'} - \alpha^j_{m'}, v_f - \beta^i_f) \geq 0 \qquad D_{ij}(u_m - \alpha^j_m, v_{f'} - \beta^i_{f'}) \geq 0$$

and consequently,

$$D_{ij}(u_m - \alpha^j_m, v_f - \beta^i_f) \leq D_{ij}(u_{m'} - \alpha^j_{m'}, v_f - \beta^i_f) \tag{32}$$
$$D_{ij}(u_{m'} - \alpha^j_{m'}, v_{f'} - \beta^i_{f'}) \leq D_{ij}(u_m - \alpha^j_m, v_{f'} - \beta^i_{f'}) \tag{33}$$

Based on Lemma 1 of GKW, $D_{ij}(u, v)$ is isotone in the sense that $(u, v) \leq (u', v')$ implies $D_{ij}(u, v) \leq D_{ij}(u', v')$ and vice-versa. Based on this property of $D_{ij}(\cdot, \cdot)$, from (35), we get $u_m - \alpha^j_m \leq u_{m'} - \alpha^j_{m'}$ and from (36), we obtain $u_m - \alpha^j_m \geq u_{m'} - \alpha^j_{m'}$. Thus, we must have

$$u_m - \alpha^j_m = u_{m'} - \alpha^j_{m'} = U_{ij} \quad \Rightarrow \quad u_m = U_{ij} + \alpha^j_m$$

By the same token, $v_f = V_{ij} + \beta^i_w$. It then follows that $D_{ij}(U_{ij}, V_{ij}) = 0$. □

Proposition 3. Given the structure for utilities in Proposition 2, individuals $m \in i$ and $f \in j$ solve the below discrete choice problems

$$u_m = \max_{j \in \{0,\ldots,J\}} U_{ij} + \alpha^j_m, \qquad v_f = \max_{i \in \{0,\ldots,I\}} V_{ij} + \beta^i_w$$

In addition, given the distribution functions $F_\alpha(\cdot)$ and $F_\beta(\cdot)$ and their corresponding density functions $f_\alpha(\cdot)$ and $f_\beta(\cdot)$, we can identify the difference between systematic parts of the utilities from the empirical matching probabilities, by solving the system of equations

$$\begin{aligned} \Pr\{m \in i,\ j = \arg\max u^k_m\} &= \Pr\{\forall k,\ \alpha^k_m \leq U_{ij} - U_{ik} + \alpha^j_m\} \\ &= \int_{-\infty}^{+\infty} \prod_{j \neq k} F_{\alpha^k}(U_{ij} - U_{ik} + \alpha^j_m)\, f_{\alpha^j}(\alpha^j_m)\, d\alpha^j_m = \frac{N_{ij}}{N_{i\oplus}} \qquad (34) \\ \Pr\{f \in j,\ i = \arg\max v^k_f\} &= \Pr\{\forall k,\ \beta^k_f \leq V_{ij} - V_{kj} + \beta^i_f\} \\ &= \int_{-\infty}^{+\infty} \prod_{i \neq k} F_{\beta^k}(V_{ij} - V_{kj} + \beta^i_f)\, f_{\beta^i}(\beta^i_f)\, d\beta^i_f = \frac{N_{ij}}{N_{\oplus j}} \qquad (35) \end{aligned}$$

With Gumbel distribution for α^j_m and β^i_f, the above equations become

$$\frac{\exp(U_{ij})}{\sum_{k=0}^{J} \exp(U_{ik})} = \frac{N_{ij}}{N_{i\oplus}} \qquad \text{and} \qquad \frac{\exp(V_{ij})}{\sum_{k=0}^{I} \exp(V_{kj})} = \frac{N_{ij}}{N_{\oplus j}} \qquad (36)$$

and we then obtain

$$U_{ij} - U_{i0} = \ln \frac{N_{ij}}{N_{i0}}, \qquad V_{ij} - V_{0j} = \ln \frac{N_{ij}}{N_{0j}} \qquad (37)$$

From Lemma 1 of GKW, the distance-to-frontier function has the following property

$$D_{ij}(a + u, a + v) = a + D_{ij}(u, v)$$

Moreover, from Proposition 2, $D_{ij}(U_{ij}, V_{ij}) = 0$, which together with (40) leads to

$$D_{ij}(U_{i0} + \ln N_{ij} - \ln N_{i0}\,,\ V_{0j} + \ln N_{ij} - \ln N_{0j}) = 0 \quad \Rightarrow \quad \ln N_{ij} = -D_{ij}(U_{i0} - \ln N_{i0}\,,\ V_{0j} - \ln N_{0j})$$

□

Proposition 4. From (6)

$$
\begin{aligned}
S_{ij} - \frac{1}{J}S_{i+} - \frac{1}{I}S_{+j} + \frac{1}{IJ}S_{++} &= S_{ij} - \frac{1}{I}\sum_{i=1}^{I} S_{ij} - \frac{1}{J}\sum_{j=1}^{J} S_{ij} + \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} S_{ij} \\
&= \frac{1}{2}\ln\frac{N_{ij}^2}{N_{i0}N_{0j}} - \frac{1}{2J}\sum_{l=1}^{J}\ln\frac{N_{il}^2}{N_{i0}N_{0l}} - \frac{1}{2I}\sum_{k=1}^{I}\ln\frac{N_{kj}^2}{N_{k0}N_{0j}} + \frac{1}{2IJ}\sum_{k=1}^{I}\sum_{l=1}^{J}\ln\frac{N_{kl}^2}{N_{k0}N_{0l}} \\
&= \ln\frac{N_{ij}\prod_{k=1}^{I}\prod_{l=1}^{J} N_{kl}^{\frac{1}{IJ}}}{\prod_{l=1}^{J} N_{il}^{\frac{1}{J}}\prod_{k=1}^{I} N_{kj}^{\frac{1}{I}}} = \ln\frac{N_{ij}N_{\times\times}}{N_{i\times}N_{\times j}} = \rho_{ij}
\end{aligned}
$$

For marriage rates, from Proposition 3, we have

$$
N_{i+} = \sum_{j=1}^{J} N_{ij} = \sum_{j=1}^{J}\exp(\ln N_{ij}) = \sum_{j=1}^{J}\exp\big(-D_{ij}(U_{i0} - \ln N_{i0},\ V_{0j} - \ln N_{0j})\big) = \sum_{j=1}^{J}\exp(S_{ij})\sqrt{N_{i0}N_{0j}}
$$

Thus, we obtain

$$
\mu_i = \frac{N_{i+}}{N_{i\oplus}} = \frac{1}{\bar{N}_{i\oplus}} = \ln\sum_{j=1}^{J}\exp(S_{ij})\sqrt{N_{i0}N_{0j}} \qquad \omega_j = \frac{N_{+j}}{N_{\oplus j}} = \frac{1}{\bar{N}_{\oplus j}} = \ln\sum_{i=1}^{I}\exp(S_{ij})\sqrt{N_{i0}N_{0j}}
$$

□

A.2 More data details

CPS is the proper data to assess inequality (of bottom 99%), compared to ACS and other data because it provides adjustment for top codings of income. We use the variables in Annual Social & Economic Supplement (ASEC) of CPS data. In addition to key variables (SERIAL, ASECWTH, ASECWT, YEAR, AGE, SEX, CPI99), the exact CPS variables are

- RELATE: Relationship to household head, available 1962-2023.
  - From 1995 forward, the "unmarried partner" code is available. Beginning in the 2019 ASEC, codes for same-sex spouses and same-sex unmarried partners are added.
- EDUC: Educational attainment recode, available 1962-2023.
  - below high school: EDUC < 72
  - high school diploma: EDUC $\in [72, 73]$
  - some college: EDUC $\in [80, 81, 90, 91, 92, 100]$
  - B.A. degree: EDUC $\in [111, 120, 110]$
  - Graduate degree: EDUC ≥ 121
- INCTOT: Total personal income, available 1962-2023.

  - indicates each respondent's total pre-tax personal income or losses from all sources for the previous calendar year. The Census Bureau applies different disclosure avoidance measures across time for individuals with high income in this variable and has provides adjustments of top income coding: `https://cps.ipums.org/cps/topcodes_tables.shtml`.
  - In CPS 1962, income is not reported for persons who were in rotation groups 4 or 8. Thus, to estimate aggregate income for this year a multiplier proportional to the weight of other rotation groups is used (variable ROTATE reports the rotation group in CPS 1962-1967).

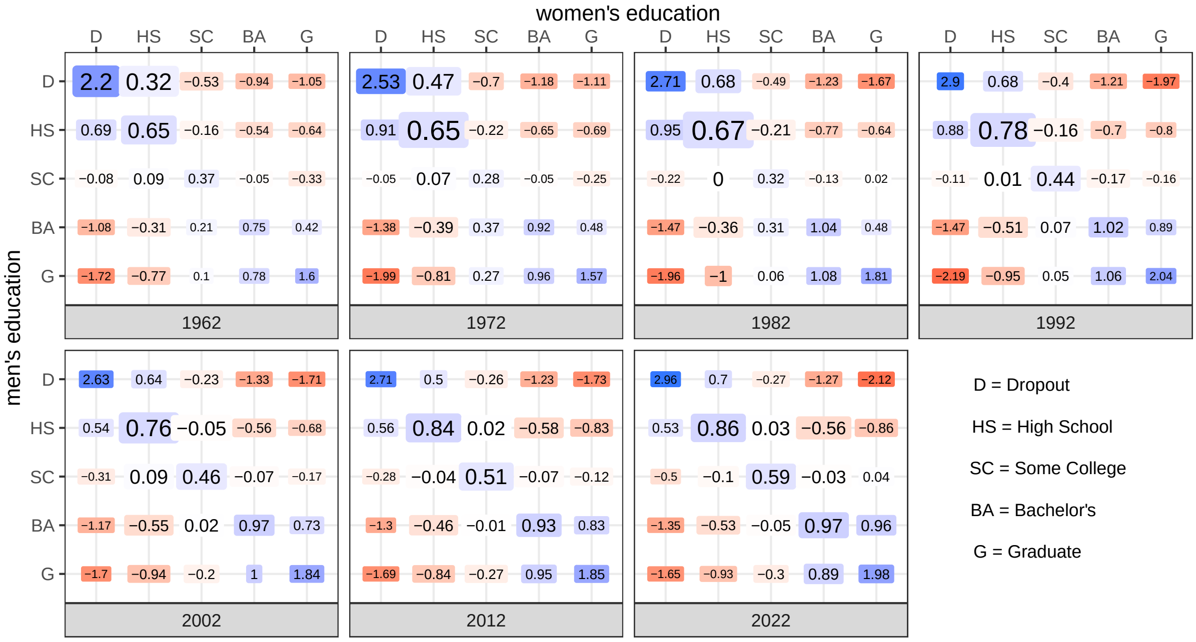
- UHRSWORKLY: Usual hours worked per week, available 1976-2023. For 1962-75, we use
  - AHRSWORKT: Hours worked last week, available 1962-2023.
- WKSWORK1: Weeks worked last year, available 1976-2023. For 1962-75, we use
  - WKSWORK2: Weeks worked last year (intervalled), available 1962-2023. To convert intervals to number of week, we compute the average of weeks by intervals for each gender and education group in 1976-1989, and use that as multiplier for this variable in 1962-75.

We compute total hours and hourly earnings as

$$\text{total hours} \;=\; \text{UHRSWORKLY} \times \text{WKSWORK2} \qquad \text{hourly earnings} \;=\; \frac{\text{INCTOT}}{\text{total hours}}$$

Table 1: Summary of population share, average household income (in 1983 dollars), and the Gini index over different marriages by education. “N.A.” corresponds to fictitious partners for singles.

men's education	women's education																		
	N.A.	D	HS	SC	BA	G	N.A.	D	HS	SC	BA	G	N.A.	D	HS	SC	BA	G	
	(1) percent of population						(2) average real income						(3) Gini index within group						
N.A.		8.08	5.21	1.43	0.84	0.53		6710	10588	13308	14142	20729		0.42	0.35	0.39	0.31	0.3	1962
D	6.52	24.49	9.21	1.55	0.43	0.11	10867	15860	20995	21695	22753	26715	0.43	0.32	0.28	0.29	0.28	0.29	
HS	3.71	4.99	11.73	1.73	0.6	0.12	15392	20708	24025	24676	31015	33669	0.32	0.26	0.24	0.28	0.31	0.18	
SC	1.46	1.12	3.29	1.61	0.55	0.09	17906	22425	28292	32109	32024	45017	0.36	0.29	0.28	0.34	0.28	0.23	
BA	0.97	0.28	2.04	1.38	1.21	0.2	22736	26778	31411	39096	38961	37472	0.34	0.25	0.26	0.35	0.33	0.22	
G	0.78	0.12	1.04	1.01	1.09	0.48	23588	31317	35484	36089	40129	43215	0.38	0.3	0.31	0.33	0.29	0.32	
N.A.		6.94	6.22	1.75	0.99	0.75		7981	12802	15890	19759	24847		0.38	0.32	0.35	0.31	0.31	1972
D	4.92	16.41	8.48	0.76	0.24	0.09	12361	20221	25916	29291	33334	35757	0.39	0.3	0.26	0.3	0.22	0.29	
HS	3.8	5.34	16.34	2.15	0.71	0.21	17705	25572	30075	32518	36124	41995	0.32	0.24	0.24	0.24	0.25	0.2	
SC	1.42	0.93	4.78	2.09	0.73	0.19	21062	28761	33217	35575	41230	40460	0.31	0.27	0.24	0.27	0.3	0.26	
BA	1.04	0.22	2.3	1.75	1.67	0.34	25914	29168	39294	42642	44672	49802	0.38	0.29	0.24	0.27	0.29	0.23	
G	1	0.1	1.36	1.39	1.62	0.98	27798	33842	46099	47773	47250	51656	0.4	0.3	0.28	0.3	0.28	0.28	
N.A.		5.49	8.56	3.78	2.18	1.71		6365	11343	13544	16558	20007		0.39	0.35	0.32	0.32	0.3	1982
D	3.98	7.87	5.38	0.65	0.13	0.04	10514	17467	22099	25001	26261	28262	0.44	0.32	0.3	0.29	0.27	0.23	
HS	6.17	3.54	14.69	2.61	0.67	0.33	15641	22227	27371	30355	31822	34915	0.37	0.28	0.25	0.26	0.26	0.22	
SC	3.33	0.72	4.96	2.96	0.93	0.4	17800	24972	30724	31469	35085	38521	0.36	0.3	0.26	0.26	0.26	0.28	
BA	2.43	0.14	2.42	2.15	2.04	0.63	22878	26740	35599	39076	39309	42011	0.36	0.3	0.26	0.27	0.26	0.25	
G	2.07	0.08	1.29	1.6	2.08	2	24921	37476	39074	43965	45218	48251	0.38	0.34	0.26	0.29	0.29	0.26	
N.A.		4.32	8.3	6.23	3.56	1.79		6412	11447	14985	20544	26976		0.42	0.38	0.35	0.34	0.31	1992
D	3.99	4.32	2.85	0.74	0.14	0.02	9174	15987	21259	24444	28437	28064	0.43	0.37	0.31	0.3	0.29	0.31	
HS	8.1	2.08	11.26	3.46	1	0.23	14068	20150	28058	31150	36199	43830	0.38	0.32	0.28	0.27	0.26	0.22	
SC	5.36	0.63	4.49	5.3	1.48	0.43	17574	24979	31984	35724	40355	44476	0.38	0.34	0.27	0.27	0.26	0.25	
BA	3.67	0.11	1.84	2.45	3.57	0.87	23211	30927	39855	44299	47097	55927	0.36	0.35	0.28	0.26	0.27	0.26	
G	1.71	0.03	0.67	1.35	1.97	1.7	30815	43481	46763	50781	55433	62733	0.38	0.29	0.28	0.24	0.26	0.26	
N.A.		3.04	6.95	6.93	4.2	1.97		7829	12580	16580	24173	35030		0.42	0.39	0.39	0.36	0.4	2002
D	3.38	3.49	2.09	0.92	0.17	0.04	9981	17627	22233	24476	34013	32135	0.4	0.37	0.34	0.34	0.42	0.22	
HS	6.96	1.62	9.15	4.21	1.47	0.47	15150	21994	30643	34792	43439	48647	0.37	0.35	0.32	0.3	0.3	0.33	
SC	5.7	0.6	4.12	6.34	2.24	0.72	19831	26046	36683	40040	46439	53086	0.41	0.34	0.32	0.3	0.3	0.3	
BA	3.98	0.18	1.61	2.93	4.82	1.49	27950	47929	46947	52664	61993	68857	0.41	0.44	0.36	0.35	0.35	0.33	
G	1.55	0.05	0.58	1.22	2.47	2.31	43132	44891	62747	63706	75876	83974	0.47	0.37	0.42	0.36	0.36	0.36	
N.A.		2.64	6.3	7.47	4.58	2.42		6359	10707	13655	22384	28971		0.38	0.41	0.4	0.41	0.39	2012
D	2.95	2.97	1.67	0.9	0.25	0.06	8989	14534	18894	23820	27737	26410	0.4	0.36	0.37	0.42	0.38	0.4	
HS	7.47	1.27	8.19	4.29	1.76	0.57	13852	18173	26237	32152	39441	43123	0.44	0.35	0.35	0.35	0.33	0.31	
SC	5.86	0.48	3.05	6.51	2.83	1.02	16670	19947	32136	34452	44689	52738	0.44	0.36	0.34	0.32	0.32	0.33	
BA	4.13	0.13	1.41	2.67	5.48	2.15	25113	23273	41285	45327	54870	64770	0.43	0.43	0.37	0.32	0.35	0.33	
G	1.56	0.04	0.47	0.97	2.6	2.87	36688	18835	48259	59041	72767	79656	0.46	0.39	0.4	0.36	0.4	0.37	
N.A.		1.69	5.95	6.89	5.38	3.15		7043	11530	14220	23784	30934		0.39	0.45	0.42	0.44	0.4	2022
D	2.3	2.26	1.31	0.57	0.21	0.04	9780	17820	20911	24877	27381	30064	0.45	0.41	0.39	0.39	0.36	0.43	
HS	8.46	0.97	7.23	3.58	2.05	0.75	13559	18911	28220	31931	41583	44596	0.42	0.36	0.37	0.34	0.36	0.32	
SC	6.01	0.28	2.34	5.44	3.03	1.52	16526	20785	33223	36487	46712	55743	0.44	0.34	0.38	0.34	0.33	0.32	
BA	5.14	0.08	1.12	2.18	6.9	3.21	26154	30866	46493	50426	61876	72005	0.44	0.46	0.41	0.37	0.36	0.33	
G	2.07	0.03	0.33	0.73	2.71	4.09	36234	30632	57852	64841	75487	86715	0.43	0.34	0.43	0.41	0.37	0.37	



1962

	D	HS	SC	BA	G
D	2.2	0.32	−0.53	−0.94	−1.05
HS	0.69	0.65	−0.16	−0.54	−0.64
SC	−0.08	0.09	0.37	−0.05	−0.33
BA	−1.08	−0.31	0.21	0.75	0.42
G	−1.72	−0.77	0.1	0.78	1.6

1972

	D	HS	SC	BA	G
D	2.53	0.47	−0.7	−1.18	−1.11
HS	0.91	0.65	−0.22	−0.65	−0.69
SC	−0.05	0.07	0.28	−0.05	−0.25
BA	−1.38	−0.39	0.37	0.92	0.48
G	−1.99	−0.81	0.27	0.96	1.57

1982

	D	HS	SC	BA	G
D	2.71	0.68	−0.49	−1.23	−1.67
HS	0.95	0.67	−0.21	−0.77	−0.64
SC	−0.22	0	0.32	−0.13	0.02
BA	−1.47	−0.36	0.31	1.04	0.48
G	−1.96	−1	0.06	1.08	1.81

1992

	D	HS	SC	BA	G
D	2.9	0.68	−0.4	−1.21	−1.97
HS	0.88	0.78	−0.16	−0.7	−0.8
SC	−0.11	0.01	0.44	−0.17	−0.16
BA	−1.47	−0.51	0.07	1.02	0.89
G	−2.19	−0.95	0.05	1.06	2.04

2002

	D	HS	SC	BA	G
D	2.63	0.64	−0.23	−1.33	−1.71
HS	0.54	0.76	−0.05	−0.56	−0.68
SC	−0.31	0.09	0.46	−0.07	−0.17
BA	−1.17	−0.55	0.02	0.97	0.73
G	−1.7	−0.94	−0.2	1	1.84

2012

	D	HS	SC	BA	G
D	2.71	0.5	−0.26	−1.23	−1.73
HS	0.56	0.84	0.02	−0.58	−0.83
SC	−0.28	−0.04	0.51	−0.07	−0.12
BA	−1.3	−0.46	−0.01	0.93	0.83
G	−1.69	−0.84	−0.27	0.95	1.85

2022

	D	HS	SC	BA	G
D	2.96	0.7	−0.27	−1.27	−2.12
HS	0.53	0.86	0.03	−0.56	−0.86
SC	−0.5	−0.1	0.59	−0.03	0.04
BA	−1.35	−0.53	−0.05	0.97	0.96
G	−1.65	−0.93	−0.3	0.89	1.98

Figure 13: AM in population measured by geometric mean of log odds ratios. The size of the boxes in the figure is proportional to the product of all four elements of the corresponding odds ratio in (1). Notably, in all years, the values of diagonal elements are positive, while the values of anti-diagonal elements are negative. Furthermore, over time, the values of the former are consistently increasing, whereas the values of the latter generally exhibit a decreasing trend. This pattern suggests a prevailing increase in assortative matching by education over time.

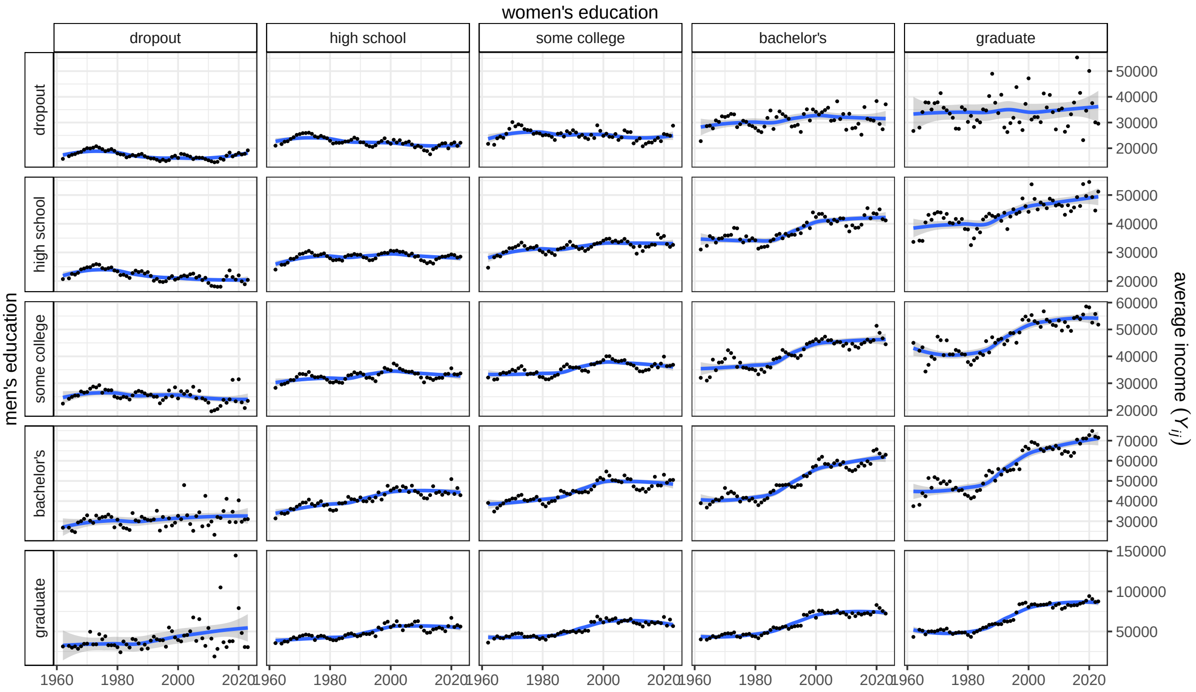


Figure 14: Average income and its LOESS estimation

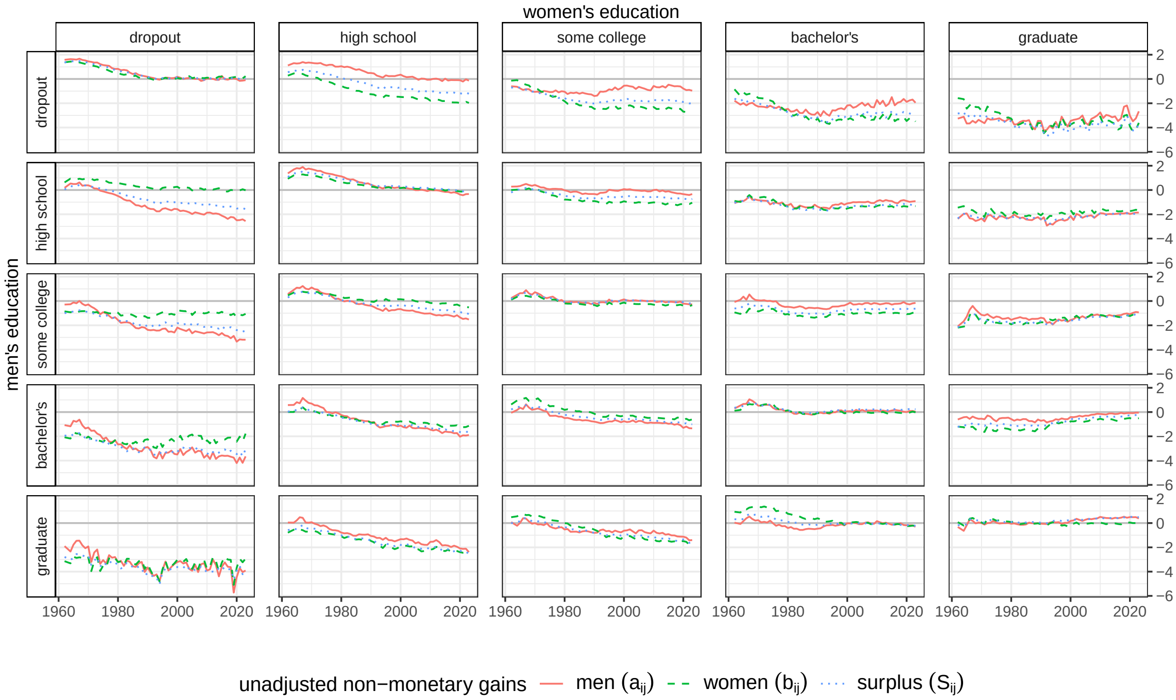


Figure 15: The estimated levels of unadjusted non-monetary gains a_{ij} and b_{ij}, and total surplus S_{ij}.

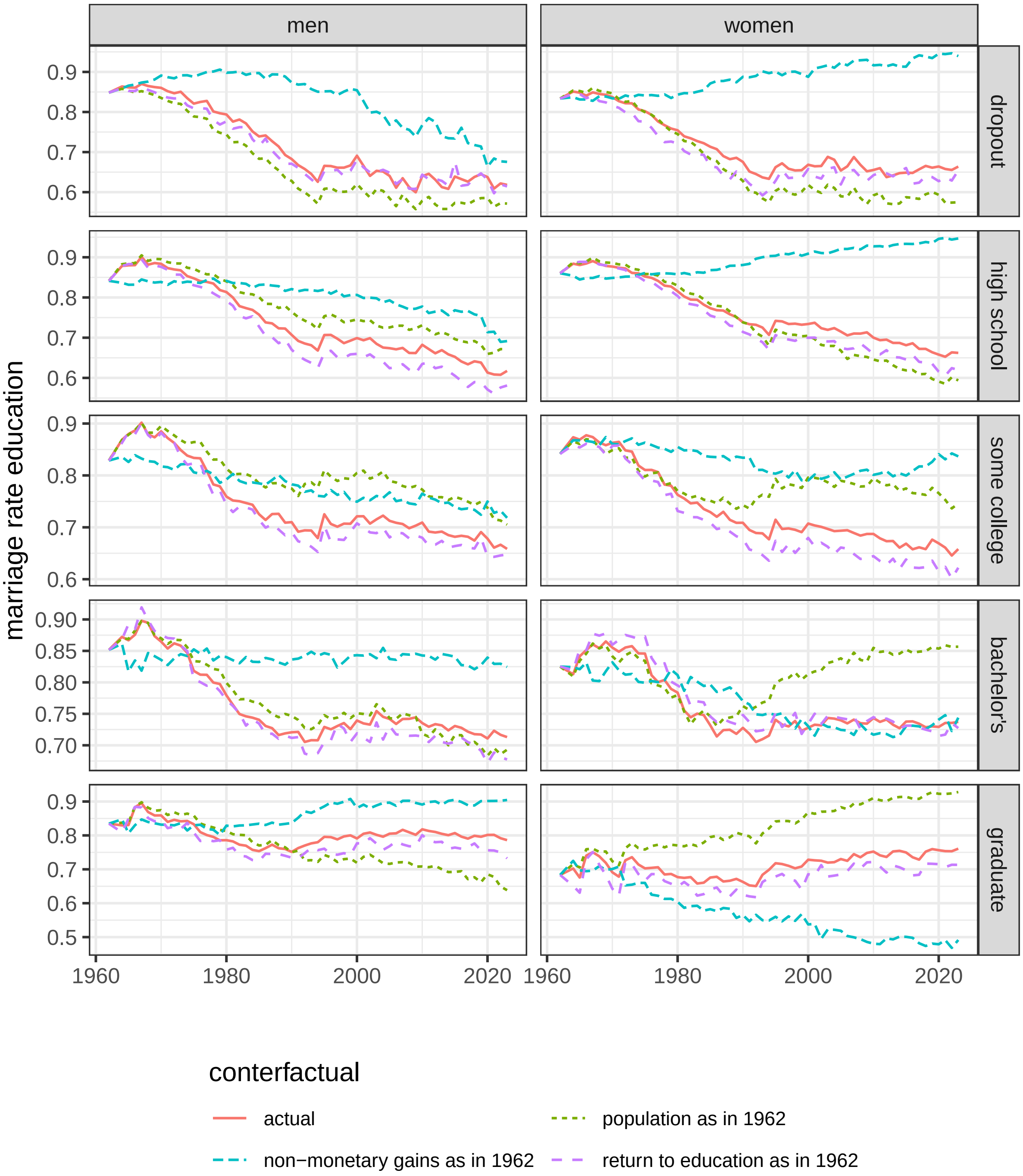


Figure 16: The counterfactual trends of marriage rate by education and gender in different scenarios.

Figure 17: The counterfactual trends of different aggregate AM index in different scenarios.

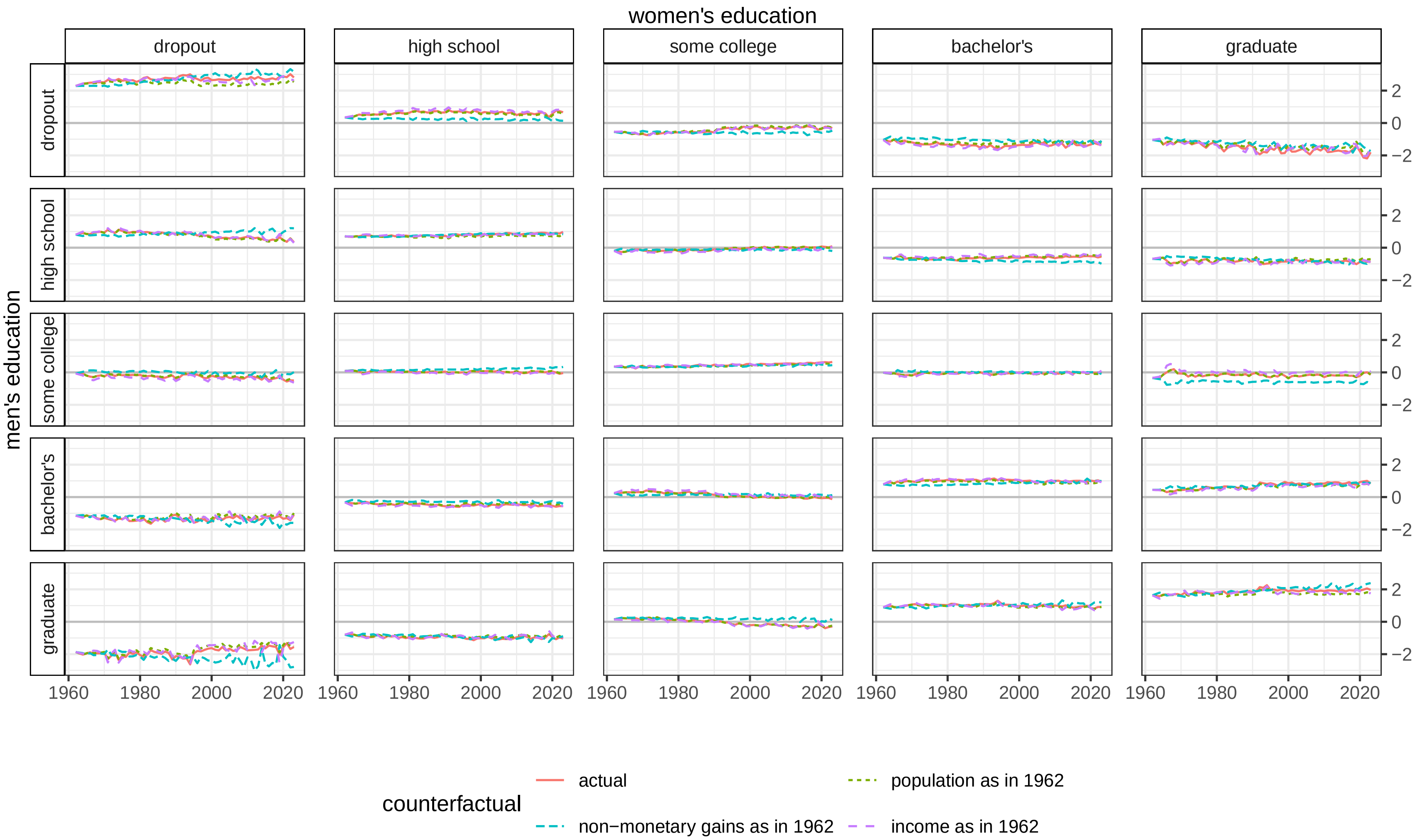


Figure 18: The counterfactual trends of AM in different scenarios.