Return to Education, Assortative Matching, and Inequality

Mohammad Hoseini*

November 29, 2024

Click here for the latest version.

Abstract

Previous studies decomposing the growth of household income inequality based on educational assortative matching (AM), assume that income distribution conditional on marriage type is independent of sorting patterns. Using a frictionless matching model with imperfectly transferable utility, we relax this assumption and account for the general equilibrium effects between the return to education (RE) and AM. The model separates AM into transferable and non-transferable components, showing that, controlling for secular RE trends, the transferable component increases inequality, while the non-transferable component reduces it. Estimation of the model using CPS data demonstrates that the rise in AM in the U.S. from 1962 to 2023 stems primarily from its non-transferable component. Consequently, after controlling for RE, AM has reduced cross-sectional income inequality. Furthermore, market returns to education are the dominant driver of inequality, explaining approximately 40% of the Gini coefficient's increase during this period. Since the tendency to invest in children's human capital is reflected in non-transferable AM, the findings suggest that highly educated couples may be so willing to spend time on their children that it reduces cross-sectional inequality but potentially intensifies long-term inequality.

JEL classifications: I24, I26, J12

Keywords: return to education, marriage market, inequality, imperfectly transferable utility

1 Introduction

Over the past century, income inequality has risen across various regions, prompting extensive research into the role of human capital investment in shaping this trend. In particular, the expansion of education has been identified as a prominent factor (e.g. Goldin and Katz, 2009; Autor, 2014). Beyond its impact on labor market outcomes, education also yields returns in the marriage market, influencing individuals' matching behavior. Evidence highlights that educational assortative matching (henceforth AM) is a key feature of marriage markets. Since the trend in AM is influenced by the return to education (henceforth

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

RE), in assessing the effect of AM on income inequality, RE serves as a confounding factor. This paper aims to address this challenge by decomposing changes in inequality into two independent components: one arising from secular trends in RE and another driven by variations in preference for AM in the marriage market.

Starting from Becker (1973, 1974), a branch of the literature investigates the role of marriage market, particularly the impact of AM on raising income inequality. A big challenge in such analysis is the measurement of sorting. Many proposed AM indices in the past literature are sensitive to changes in the marginal distribution of the population, which can dramatically change over time and unevenly across genders. Consequently, these indices often capture a combination of sorting and population trends rather than preferences for AM (Chiappori, Costa Dias, and Meghir, 2021). Recent studies in this field (e.g. Eika, Mogstad, and Zafar, 2019; Chiappori, Costa-Dias, Crossman, and Meghir, 2020; Dupuy and Weber, 2022) address this problem by employing AM indices that are independent of the marginal distribution of the population. This approach allows them to let the education distributions evolve according to their actual trends while fixing AM at a benchmark year. By doing so, they can estimate the contribution of changes in AM to the growth of income inequality.

To estimate the share of AM in rising inequality, these studies adopt the standard inequality decomposition framework (Fortin, Lemieux, and Firpo, 2011). This approach relies on the assumption of conditional independence, which posits that the conditional income distribution for each type of couple remains fixed when the sorting pattern changes in the counterfactual scenario. In other words, it assumes that while any gains from marriage in terms of household income remain at their current levels, households marry and sort according to the pattern observed in the base year. However, this assumption disregards the general equilibrium effects of trends in RE on marriage market outcomes. For instance, as macroeconomic factors change RE, the 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 this paper, we employ a matching model to relax the conditional independence assumption used in previous studies. This allows us to disentangle the effects of secular trends in RE when estimating the contribution of AM to household income inequality, and vice versa. The model is a frictionless matching framework with imperfectly transferable utility, as developed by Galichon, Kominers, and Weber (2019) (hereafter GKW). This approach links household formation to the allocation of power within households, both of which are jointly determined in the marriage market equilibrium.

The main result of the theoretical model is that marriage market outcomes, including marriage rates and AM, depend on the population of singles and the marriage surplus. The marriage surplus is defined as the joint gain from marriage minus the sum of the gains each individual would achieve if they remained single. This surplus comprises two components: one derived from non-transferable gains, which are independent of income, and another from transferable gains that are linked to income through individual's consumption. While the non-transferable component arises from unobservable factors in the marriage (e.g., children, love), the sources of transferable component are one's own income and his/her spouse's income. Both components can be identified using contingency tables of population and average income, combined with an assumption about the income-sharing rule within each couple type and the degree of utility transferability.

Based on the model, any change in AM arises from the supermodularity of either the non-transferable or transferable components of the marriage surplus. Furthermore, in line with approaches that separate AM from population marginals, we define RE indices corresponding to the secular trends in non-pecuniary and pecuniary gains of education. These indices are constructed to be orthogonal to changes in the two components of AM. They capture the expected per capita non-pecuniary and pecuniary gains associated with different educational levels at a given time which can be reasonably assumed to be independent of matching patterns. By adopting this framework, we can isolate the impact of AM on income inequality while controlling for the mechanical effects of RE changes on inequality.

The theory demonstrates that, controlling for RE, changes in AM driven by its transferable component generally increase household income inequality. In contrast, an increase in the non-transferable component of AM, holding other factors constant, tends to reduce cross-sectional inequality. This latter finding contrasts with the conventional notion that increased AM widens the household income distribution by concentrating more high-earning or low-earning partners in the same households. The intuition behind this finding is as follows: Suppose individuals' utility is the sum of a non-transferable component (e.g., affinity, children) and a *concave* function of a transferable good (income), and the total expected gains from education in the economy are fixed, regardless of the matching pattern. In this setting, if preferences for sorting arise from the non-transferable component rather than income, increased AM redistributes a larger share of total income to couples where both partners are less educated, reducing income inequality. Conversely, if sorting preferences are driven by the transferable component of utility, higher AM results in a greater share of income for couples where both partners are highly educated, thereby increasing household income inequality.

We estimate counterfactual income inequality for the United States 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. However, nearly all of this increase is attributable to the non-transferable component of AM, with its transferable term remaining almost randomly matched throughout the years. When controlling for RE, our counterfactual analysis reveals a significant increase in the Gini coefficient in 2023 if AM components are fixed at their 1962 levels, consistent with our theoretical predictions. This finding is robust across different assumptions about the degree of transferability and the income-sharing rule. Additionally, we find a substantial reduction in income inequality when the pecuniary component of RE is fixed at its 1962 levels. Under this scenario, the Gini coefficient decreases by 4 points for both married couples and all households, accounting for approximately 40 percent of the overall rise in income inequality between 1962 and 2023.

This paper contributes to the literature on inequality from a household economics perspective. Since Becker (1973, 1974), AM has been a central focus in studies examining household income inequality within the marriage market. Although significant increases in AM have been documented in many countries, empirical studies decomposing cross-sectional household inequality have often found its impact to be negligible (Greenwood, Guner, Kocharkov, and Santos, 2015; Eika et al., 2019; Chiappori et al., 2020; Ciscato and Weber, 2020; Dupuy and Weber, 2022). A critical limitation of these analyses is their failure to account for the general equilibrium effects between AM and RE when assessing the relationship between AM and inequality, despite theoretical arguments suggesting that AM is positively correlated with the market return to human capital (Fernandez, Guner, and Knowles, 2005; Chiappori, Salanié, and Weiss, 2017). Standard decomposition methods typically impose an identification constraint that assumes the income distribution, conditional on education, remains independent of changes in AM in counterfactual scenarios (Fortin et al., 2011). This paper relaxes that assumption by "opening the black box" of AM and decomposing it into non-transferable and transferable components.

A further methodological contribution of this paper is to endogenize the decision to marry by linking it to RE. In most previous studies, singles are either excluded from the decomposition or included in the model with trends assumed to be exogenous to educational dynamics. By addressing this limitation, this paper provides a more comprehensive framework for analyzing the role of the marriage market in shaping income inequality.

Our study challenges the conventional notion that when individuals match assortatively, their combined earnings diverge more, thereby increasing cross-sectional income inequality compared to random matching. Due to the concavity of pecuniary gains from education, the model implies that when AM arises because of non-transferable factors, household income by marital type adjusts in a way that reduces household income inequality. This finding suggests that marriage can play a role in mitigating current household income inequality, although its effect on persistent inequality can be in the opposite direction. This is because other mechanisms, such as investments in children and intergenerational wealth transfers, are not captured in cross-sectional inequality measures.

To clarify, the analysis in this paper focuses on inequality within a cross-section of the population rather than across generations. From a static perspective, AM directly influences inequality across households. Contrary to the conventional view, this paper argues that the direct effect of AM on inequality can be decreasing when RE is controlled for, a prediction supported by the U.S. data. From a dynamic perspective, however, AM has long-term implications for intergenerational mobility, as educated couples tend to invest more in their children's human capital. While this paper does not explicitly model this mechanism, it acknowledges its importance, particularly as children are a key element of non-transferable marital gains.

The empirical finding of the significant rise in non-transferable AM in the U.S. highlights the growing importance of the production of children's human capital, which requires input from both parents. As Chiappori et al. (2017) argue, when the return to human capital investment increases, high-income couples devote more resources—particularly time—to their children, reinforcing AM. In our model, this trend is reflected in the non-transferable component of AM. Thus, our findings are consistent with this mechanism and further suggest that the tendency for investing in children may be so substantial that it leads to a reduction in cross-sectional household income inequality.

The rest of the paper is organized as follows: Next section provides a simple example showing how relaxing conditional independence assumption might change the effect of AM on inequality. Section 3 outlines the measurement of AM based on the association of row and columns of matching tables and reviews the standard inequality decomposition method. Section 4 develops the matching model and discusses its identification. Section 5 analyzes the relationship between RE, AM, and inequality and describes the procedures to build various counterfactual scenarios. Section 6 describes the data, overall trends, and estimated parameters and section 7 presents the counterfactual trends of the U.S. income inequality. Finally, section 8 concludes.

2 An Illustrative Example

Consider a marriage market with an equal population of men and women, where individuals are classified based on their level of human capital, and all participants are matched with a partner. Let i and j denote the indices of education level for men and women, respectively. Define N_{ij} as the population, Y_{ij} as the household income, and λ_{ij} and $1-\lambda_{ij}$ as the share of income consumed by male and female partners of couple type ij, respectively. Assume that the aggregate income $\sum_i \sum_j N_{ij} Y_{ij} = C$, which equals total household consumption expenditure, is determined exogenously by factors outside the marriage market and remains independent of the matching pattern.

Within this framework, we define the secular trends of RE as the expected total consumption expenditure of individuals based on their education level, irrespective of their matching type. Let $N_{i+} = \sum_i N_{ij}$ and $N_{+j} = \sum_j N_{ij}$ represent the total population of men with education i and women with education j, respectively. Similarly, let $C_{i+} = \sum_i N_{ij}\lambda_{ij}Y_{ij}$ and $C_{+j} = \sum_j N_{ij}(1-\lambda_{ij})Y_{ij}$ be total consumption expenditure by men with education i and women with education j. Then, we define RE indices for men with education i and women with education j by C_{i+}/N_{i+} and C_{+j}/N_{+j}, respectively. These indices capture the expected pecuniary returns to education, combining income from one's own labor market outcomes and potential gains from a spouse's income in marriage. In other words, they measure the total expected pecuniary gains from education, encompassing both labor market and marriage market outcomes.

To simplify the analysis in the rest of this section, we assume there are two levels of education $i = \{1,2\}$, $j = \{1,2\}$ and $\forall i,j:\ N_{i+} = N_{+j} = 2$, $\lambda_{ij} = 0.5$. Moreover, $C_{1+} = C_{+1} = C_1$ and

$C_{2+} = C_{+2} = C_2$. Our aim is to compare income inequality under two matching scenarios:

- Assortative matching: $N_{11} = N_{22} = 2$ and $N_{12} = N_{21} = 0$,
- Random matching: $N_{11} = N_{12} = N_{21} = N_{22} = 1$,

based on two assumptions about households' income:

(i) Y_{ij} is independent of who marries whom.

(ii) Y_{ij} depends on matching patterns, but RE indices $\frac{1}{2}C_{i+}$ and $\frac{1}{2}C_{+j}$ are independent of that.

We further assume that income is increasing in education such that under any matching scenario and either of the above assumptions, we have $Y_{11} \leq \{Y_{12}, Y_{21}\} \leq Y_{22}$. To measure household inequality, we compute the total income transfer T required from above average households to below average ones in order to have perfect equality such that $\forall i, j: \; Y_{ij} = \bar{Y}$.

In this setting, under assumption (i), assortative matching always leads to higher household income inequality as shown in Figure 1. Under assumption (ii), $\bar{Y} = \frac{1}{2}(C_1 + C_2)$ and, while assortative matching leads to $Y_{11} = C_1$ and $Y_{22} = C_2$, random matching results in

$$Y_{12} = Y_{21} = 2C_1 - Y_{11}, \quad Y_{22} = 2(C_2 - C_1) + Y_{11}$$

Thus, under random matching the income distribution depends on the level of Y_{11} (any other element of income table can be the benchmark too). In this case the difference in inequality of the two scenarios become

$$T^{\mathrm{RM}} - T^{\mathrm{AM}} = \frac{1}{2}(C_2 - 3C_1) + Y_{11}$$

In this setting, assortative matching generates higher inequality only when $0 \leq Y_{11} < \frac{1}{2}(3C_1 - C_2)$, and when $C_2 > 3C_1$ or $Y_{11} > \frac{1}{2}(3C_1 - C_2)$, inequality is higher in the random matching case.

This example illustrates how the relationship between AM and inequality depends on the assumption regarding the independence of income by type from the matching pattern. In the subsequent sections, we employ a matching model to establish a link between the income table and the population table, while assuming that RE indices, similar to those defined above, are exogenously given. Before delving into the model, we first describe the issues related to the proper measurement of sorting and decomposition methods for income inequality, which are essential prerequisites for constructing counterfactual experiments.

3 Inequality Decomposition by Educational Sorting

In this section, we first set out the measurement of AM such that it is independent to changes in marginal distribution of population. We then describe the standard decomposition practice to assess the contribution of AM and its underlying assumptions.

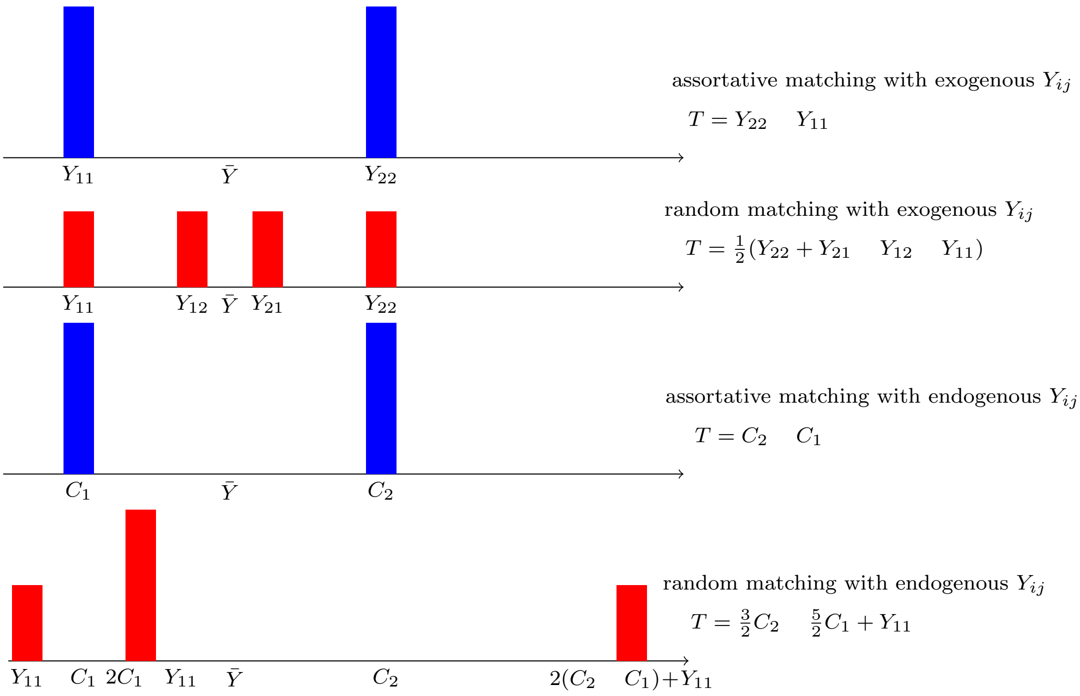

Figure 1: Income distribution under two matching scenario and two assumptions on independence of household income and matching patterns.

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 $i \in \{1, \ldots, I\}$, and the columns correspond to women with education levels $j \in \{1, \ldots, J\}$. 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 subscript to denote summation starting from 0 and 1, respectively. Thus, N_{i+} represents the population of married men with education level i, and $N_{i} \; = N_{i0} + N_{i+}$ represents the total population of men with education level i. Similarly, N_{+j} and $N \;_{j}$ represent the populations of married women and all women with education level j, respectively. In the rest of the paper, we use normal font for elements (e.g. $N_{ij}, N_{i0}, N_{i+}, N_{i} \;$) and bold fonts for vectors and matrices (e.g. $\boldsymbol{N}, \boldsymbol{N}_{0}, \boldsymbol{N}_{+}, \boldsymbol{N} \;$).

Using this notation, we define marriage rates for men with education i and women with education j as $\mu_i = N_{i+}/N_{i} \;$ and $\omega_j = N_{+j}/N \;_{j}$, 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 AM, which is an intensive margin index capturing spouse quality by education, conditional on marriage.

3.1 Measurement of educational assortative matching

Measuring sorting 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.[1]

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

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

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

where $\bar{N}_{i} \ = \prod_{j=1}^{J} N_{ij}^{1/J}$, $\bar{N}_{\ j} = \prod_{i=1}^{I} N_{ij}^{1/I}$ and $\bar{N} \ = \prod_{i=1}^{I}\prod_{j=1}^{J} N_{ij}^{1/(IJ)}$ 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 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.

3.1.1 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:

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

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

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):

$$\rho = \sum_{i=1}^{I} \sum_{j=1}^{J} \rho_{ij} \sum_{t=1}^{T} \frac{N_{ij}^t \; \bar{N}_i^t \quad \bar{N}^t{}_{\;j}}{T \sum_k \sum_l N_{kl}^t \; \bar{N}_k^t \quad \bar{N}^t{}_{\;l}} \tag{3}$$

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

3.2 Building matching table by population vectors, marriage rates, and AM matrix

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. Here, 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. Intuitively, given that in one-to-one matching $\sum_i N_{i+} = \sum_j N_{+j}$, the marriage rate combined with marginal distributions provide $I + J - 1$ 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*

- *educational distribution vectors $\boldsymbol{N}$ and $\boldsymbol{N}$,*
- *marriage rate vectors $\boldsymbol{\mu}$ and $\boldsymbol{\omega}$, such that $\sum_{i=1}^{I} \mu_i N_i \; = \sum_{j=1}^{J} \omega_j N \;_j$, and*
- *educational assortative matching matrix $\boldsymbol{\rho}$ or any other 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 from marginal population vectors, marriage rate vectors, and the AM matrix. This means that we can make counterfactual exercise by fixing any of these component at a benchmark year and find the equilibrium matching table.

3.3 Decomposition of income inequality

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

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

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

$$\widehat{F}_Y(y|\, t) = \int F_{Y|X}(y|\, x,t)\, \Psi(x|t,t_x) dF_X(x|\, t), \qquad \Psi\big(x|\, t,t_x\big) = \frac{d\hat{F}_X(x|\, t_x)}{dF_X(x|\, t)}$$

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 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:

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

Here, $F_{Y|I,J}(y|\, i,j,t)$ 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 several studies (Eika et al., 2019; Chiappori et al., 2020; Ciscato and Weber, 2020; Dupuy and Weber, 2022) for different countries. Despite different measures of AM suggesting an increasing trend, the counterfactual trend for constant AM is found to have a negligible difference with the actual inequality trend. One reason for this result could be the assumption of invariant conditional distribution of income over time which assumes $F_{Y|X}(y|\,x,t)$ 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 income table on AM. Essentially, it supposes that, while the gains of marriages, which generally depends on RE, is changing over time, households sort in the same pattern as the base year. However, during periods when economic factors significantly influence income by education, the economic gains of marrying a partner with different human capital undergo uneven changes. Consequently, the absence of a connection between average income by couple type and AM fails to capture variations in the incentive structure for marital sorting resulting from exogenous economic 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 household survey belonging to education groups (i,j), we assume that the counterfactual income becomes

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

where $\widehat{Y}_{ij}$ is the adjusted average income of couples in educational groups ij after accounting for the impact of population changes, secular trends in RE, and the AM pattern. We can then compute a counterfactual inequality index like Gini coefficient by rewighting the sample multiplier for household (m,w) 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 population matrix with elements $\widehat{N}_{ij}$, we need to construct income matrix with elements $\widehat{Y}_{ij}$ representing the predicted average income within each matched group in the counterfactual scenario. These parameters are the outcome of equilibrium in the marriage market, and to find them we need a matching model that accounts for both pecuniary and non-pecuniary gains of marriage. In the next section, we provide a frictionless matching model with imperfectly transferable utility to characterize the average income table based on return to education and the marriage market outcomes.

4 Matching Model

In this section, we present the theoretical framework that we later use for building 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 education, which influences both labor income and marital gains, on matching decisions. At the household level, non-pecuniary gains from marriage are exogenous to household income and non-transferable. In contrast, pecuniary gains depend on the state of the economy, particularly the market return to human capital, and these gains are imperfectly transferable between partners through consumption.

The main result of the model is that marriage market outcomes, including marriage rates and AM, are functions of the population 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-transferable gains and the other to transferable gains from marriage. Both components can be identified using the contingency tables of population and average income, along with an assumption about the transferability parameter and the sharing rule within each couple type.

4.1 Imperfectly transferable utility

Suppose 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.[2] A proper bargaining set has a corresponding distance-to-frontier function defined by

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

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 $\nu_{mf} = 1$,
- $u_m \geq u_{m0}$ 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.

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

4.2 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,[3] 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 $v_f = V_{ij} + \beta^i_f$.*

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.*

[3] 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.

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,*

$$\rho_{ij} = S_{ij} - \frac{1}{I}\sum_{i=1}^{I} S_{ij} - \frac{1}{J}\sum_{j=1}^{J} S_{ij} + \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} S_{ij}$$
$$\mu_i = \frac{1}{N_i}\sum_{j=1}^{J} \exp(S_{ij})\sqrt{N_{0j}N_{i0}}$$
$$\omega_j = \frac{1}{N\ _j}\sum_{i=1}^{I} \exp(S_{ij})\sqrt{N_{i0}N_{0j}}$$

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.

4.3 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 $c_i + c_j \le Y_{ij}$, 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

$$\mathcal{B}_{ij} = \left\{ (u,v) \in \mathsf{R}^2,\ \mathcal{U}_i^{\ 1}(u) + \mathcal{V}_j^{\ 1}(v) \le Y_{ij} \right\}$$

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

In the collective framework, household solves

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

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.

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

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.[4] Few classes of bargaining sets, including the ones explored in GKW, have closed-form distance functions. The most 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

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

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.

4.4 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 non-transferable marital gains for men and women, respectively. These terms include both public goods such as children and non-economic gains from marriage such as love and companionship. 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 $\gamma_{ij} = \bar{\gamma} + \epsilon_{ij}$, where ϵ_{ij} has a mean of zero and finite variance conditional on $\bar{\gamma}$, such that as $\bar{\gamma}$ 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

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

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 non-transferable marital gains ($b_{ij} - a_{ij}$), where the partner with lower non-transferable 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

$$S_{ij} = \gamma_{ij} \ln Y_{ij} - \gamma_{ij} \ln \Big(Y_{i0} \big(\frac{N_{0j}}{N_{i0}}\big)^{\frac{1}{2\gamma_{ij}}} \exp(-\frac{a_{ij}}{\gamma_{ij}}) + Y_{0j} \big(\frac{N_{i0}}{N_{0j}}\big)^{\frac{1}{2\gamma_{ij}}} \exp(-\frac{b_{ij}}{\gamma_{ij}}) \Big) \tag{12}$$

which suggests that the surplus increases with non-transferable gains as well as household income Y_{ij}, while it decreases with income if remaining single Y_{i0} and Y_{0j}. However, the impact of single population ratio N_{i0}/N_{0j} can go in either way. Using (11), we can also find the surplus as a function of Pareto weight as

$$S_{ij} = \underbrace{\frac{1}{2}(a_{ij} + b_{ij})}_{\substack{\text{non-transferable} \\ \text{component}}} + \underbrace{\gamma_{ij} \ln \Big(Y_{ij} \sqrt{\frac{\lambda_{ij}(1-\lambda_{ij})}{Y_{i0} Y_{0j}}} \Big)}_{\text{transferable component}} \tag{13}$$

Therefore, for given values of average non-transferable gains and income by type, the surplus is maximized when an even Pareto weight is applied within households. Furthermore, one can show that $c_i = \lambda_{ij} Y_{ij}$ and $c_j = (1 - \lambda_{ij}) Y_{ij}$. Thus, the Pareto weight is also the private consumption sharing rule in the model which is not directly affect the non-transferable component of utilities. Another implication of equation (13) is the decomposition of the marriage surplus into two components: a non-transferable component, which is independent of income, and a transferable component, which is determined by the market return to education and sharing rule. Later, we will use this decomposition in our counterfactual exercises.

4.5 Parameter identification

The above model has three parameter sets to identify for each couple type: $\boldsymbol{a}, \boldsymbol{b}$, and $\boldsymbol{\gamma}$. To identify martial gains, note that ETU model leads to this matching function for each couple type

$$N_{ij} = \left(\frac{Y_{i0}}{Y_{ij}} \left(N_{i0} e^{a_{ij}} \right)^{\frac{-1}{\gamma_{ij}}} + \frac{Y_{0j}}{Y_{ij}} \left(N_{0j} e^{b_{ij}} \right)^{\frac{-1}{\gamma_{ij}}} \right)^{\gamma_{ij}} \tag{14}$$

and by combining (11) and (14), we obtain

$$a_{ij} = \gamma_{ij} \ln \frac{Y_{i0}}{\lambda_{ij} Y_{ij}} + \ln \frac{N_{ij}}{N_{i0}}, \qquad b_{ij} = \gamma_{ij} \ln \frac{Y_{0j}}{(1-\lambda_{ij}) Y_{ij}} + \ln \frac{N_{ij}}{N_{0j}} \tag{15}$$

Thus, upon having information on sharing rule $\boldsymbol{\lambda}$, marital gain parameters can be readily identified from (15), given the table of $\boldsymbol{\gamma}$. Still, we can identify the below lower-bounds for the marital gains

$$a_{ij} \geq \gamma_{ij} \ln \frac{Y_{i0}}{Y_{ij}} + \ln \frac{N_{ij}}{N_{i0}}, \qquad b_{ij} \geq \gamma_{ij} \ln \frac{Y_{0j}}{Y_{ij}} + \ln \frac{N_{ij}}{N_{0j}} \tag{16}$$

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}. Thus, when information on these transfer are unavailable, one would need additional assumptions to identify these parameters.

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

$$\lambda_{ij} \in \begin{cases} (\underline{\lambda}_{ij}, 1] & \text{if } N_{i0} \exp(a_{ij}) < N_{0j} \exp(b_{ij}) \\ \underline{\lambda}_{ij} & \text{if } N_{i0} \exp(a_{ij}) = N_{0j} \exp(b_{ij}) \\ [0, \underline{\lambda}_{ij}) & \text{if } N_{i0} \exp(a_{ij}) > N_{0j} \exp(b_{ij}) \end{cases} \qquad \text{where } \underline{\lambda}_{ij} = \frac{Y_{i0}}{Y_{i0} + Y_{0j}} \tag{17}$$

Under the non-transferable utility (NTU) case, where $\gamma_{ij} \to 0$, the Pareto weight can be 0, $\underline{\lambda}_{ij}$, or 1, depending on the comparison between $N_{i0} \exp(a_{ij})$ and $N_{0j} \exp(b_{ij})$. In the TU case, where $\gamma_{ij} \to +\infty$, $\underline{\lambda}_{ij}$ is the only possible Pareto weight, regardless of the direction of the inequality in the condition. Therefore, a reasonable choice for the Pareto weight in absence of external data and regardless of γ_{ij}, is $\underline{\lambda}_{ij}$. When $\lambda_{ij} = \underline{\lambda}_{ij}$, couples allocate their household income based on their potential income if they remained single. Then, in equilibrium, the population ratio of singles will reflect the ratio of the non-pecuniary gains that are not transferable. Under this scenario, we also obtain

$$S_{ij} = \frac{1}{2}(a_{ij} + b_{ij}) + \gamma_{ij} \ln \frac{Y_{ij}}{Y_{i0} + Y_{0j}} \tag{18}$$

which suggests that the transferable part of the surplus is equal to the log of the ratio of couple income to the sum of single's income.[5]

According to Theorem 5 of GKW, identification of γ_{ij} requires information on transfer across multiple markets. In case of data limitations on multi-market transfers, one might calibrate γ_{ij} by leveraging the homoskedasticity of random terms in utilities such that the transferable component of the household utility also becomes homoskedastic. According to Proposition 2 and Assumption 2, the stochastic part of the surplus is $\frac{1}{2}(\alpha_m^j - \alpha_m^0 + \beta_f^i - \beta_f^0)$ which is the average of two standard logistic random variables. Assuming log-normal distribution for the income of each couple type, one can choose γ_{ij} such that the variance of the transferable term of the surplus in (18) when $\lambda_{ij} = \lambda_{ij}$ equals $\pi^2/3$ which is the variance of a standard logistic random variables. This yields

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

5 Marriage Market and Inequality

In this section, we describe how our matching model can be used to analyze the relationship between marriage market outcomes and overall income inequality. We begin by defining the secular trends of RE in a way that ensures its evolution is independent of matching patterns, similar to the marginal distribution of the population in Proposition 1. Next, we demonstrate how the model allows us to decompose AM into its non-transferable and transferable components. To provide theoretical insights, we then examine a simplified case with two education levels, analyzing the impact of various factors on overall inequality. Finally, we outline the algorithm used to compute equilibrium in the general framework.

5.1 Pecuniary and non-pecuniary secular trends of RE

Education generates returns in both the labor and marriage markets. While labor market returns are solely pecuniary, marriage market returns include both pecuniary and non-pecuniary components. The pecuniary benefits in marriage arise from higher consumption enabled by spousal income, whereas the non-pecuniary benefits includes factors such as love, companionship, and children. Consequently, the source of non-pecuniary return to education is only the marriage market, but the pecuniary return are derived from both labor and marriage markets. For instance, in a household where the husband works and the wife is a homemaker, if income is evenly shared, the overall pecuniary return to education is the same for both partners, even though the husband is the primary earner. In this case, the pecuniary return to education is determined by the household's sharing rule rather than by which partner earns the income.

[5] Note that the use of λ_{ij}^* and the equality of $e^a N_{i0} = e^b N_{0j}$ is only for the finding of a_{ij} and b_{ij} in each year and it does not mean that in building counterfactual scenarios λ_{ij} is constant. In estimations, we try to use alternative levels of λ_{ij} to identify model parameters as robustness analysis.

Therefore, given the sharing rule for each couple type, we can define

$$C_{i} \;\; = N_{i0}Y_{i0} + \sum_{j=1}^{J} N_{ij}\lambda_{ij}Y_{ij}, \qquad C_{\;\;j} = N_{0j}Y_{0j} + \sum_{i=1}^{I} N_{ij}(1-\lambda_{ij})Y_{ij} \tag{20}$$

where C_{i} and $C_{\;\;j}$ represent the total consumption expenditure by men with education level i and women with education level j, respectively, in the entire economy. Importantly, the source of income is irrelevant in this context and these variables capture the total household spending associated with individuals of specific gender and education levels. The crucial factor here is the decision-making power within the household, not which partner generates the income.

Similar to how the vectors $\boldsymbol{N}$ and $\boldsymbol{N}$ reflect the marginal distributions of population, the vectors $\boldsymbol{C}$ and $\boldsymbol{C}$ represent the marginal distributions of total household expenditure attributable to individuals of specific gender and education levels, independent of the matching patterns in the marriage market. A variety of factors outside the marriage market shape these expenditure distribution vectors $\boldsymbol{C}$ and $\boldsymbol{C}$. In additions to economic variables such as population supplies, labor demand, and labor productivity by human capital, these vectors are determined by non-economic factors such as social norms regarding gender roles, legal frameworks for marriage, and the balancing effect of higher education on gender parity.

In this context, we define the *pecuniary RE* index by gender as

$$\phi_i^m = \frac{C_{i}}{N_{i}}, \qquad \phi_j^f = \frac{C_{\;\;j}}{N_{\;\;j}} \tag{21}$$

which is determined by the secular trends in population and household expenditure, independent of the marriage market. This definition of return to education includes both economic gains from one's own labor income and also his/her spousal income, effectively combining the labor return and the economic component of the marriage return to education. Another advantage of this definition is its monetary unit which is the unit of the single's income. In addition, we can simply measure relative return by computing the ratios. For instance, male's pecuniary return to education i relative to i' is $\phi_i^m/\phi_{i'}^m$.

In contrast, education creates non-pecuniary return through marriage which is reflected in parameters $\boldsymbol{a}$ and $\boldsymbol{b}$ in the above model. To measure this return in the unit of pecuniary returns, we let

$$D_{i} \;\; = N_{i0} + \sum_{j=1}^{J} N_{ij}\exp(\frac{a_{ij}}{\gamma_{ij}}), \qquad D_{\;\;j} = N_{0j} + \sum_{i=1}^{I} N_{ij}\exp(\frac{b_{ij}}{\gamma_{ij}}) \tag{22}$$

and define *non-pecuniary RE index* by gender as[6]

$$\varphi_i^m = \frac{D_{i}}{N_{i}}, \qquad \varphi_j^f = \frac{D_{\;\;j}}{N_{\;\;j}} \tag{23}$$

[6]Note that from (22), we must have $D_{i\oplus} \quad N_{i+}$ and $D_{\oplus j} \quad N_{+j}$.

φ_i^m (φ_j^f) show the expected non-pecuniary gain from marriage for men (women) with education i (j) measured in the unit of private consumption. Similar to above, the relative non-pecuniary return for each gender can be defined as $\varphi_i^m/\varphi_{i'}^m$ and $\varphi_j^f/\varphi_{j'}^f$, too.

These definitions allow us to decompose the contribution of AM and marriage rates on income inequality, while controlling for the effect of RE as defined by the expected gain of an individual with a given level of education in the unit of single's income.

5.2 Transferable and non-transferable components of AM

Similar to return to education, AM can be decomposed into transferable and non-transferable components. Let $\Delta(\cdot)$ be the demean operator defined as $\Delta(x_{ij}) = x_{ij} - \frac{1}{J}x_{i+} - \frac{1}{I}x_{+j} + \frac{1}{IJ}x_{++}$. Then, according to Proposition 4, AM measured by ρ_{ij} is the demean of surplus. Because $\Delta(\cdot)$ is a linear operator, by defining

$$\rho_{ij}^a = \Delta(a_{ij}), \quad \rho_{ij}^b = \Delta(b_{ij}), \quad \rho_{ij}^Y = \Delta\Big(\gamma_{ij}\ln\big(Y_{ij}\sqrt{\frac{\lambda_{ij}(1-\lambda_{ij})}{Y_{i0}Y_{0j}}}\big)\Big) \tag{24}$$

from (13), we obtain $\rho_{ij} = \frac{1}{2}(\rho_{ij}^a + \rho_{ij}^b) + \rho_{ij}^Y$.

Therefore, AM can be decomposed into non-transferable and transferable components, defined by $\frac{1}{2}(\rho_{ij}^a + \rho_{ij}^b)$ and ρ_{ij}^Y, respectively. The non-transferable component is the average of the demeaned non-pecuniary gains of the two partners. Moreover, from the model, $c_{ij}^m = \lambda_{ij}Y_{ij}$ and $c_{ij}^f = (1-\lambda_{ij})Y_{ij}$ are the consumption expenditure of the husband and wife, respectively, in a couple of type ij, respectively. Then, if $\forall i,j:\ \gamma_{ij}=\gamma$,

$$\rho_{ij}^Y = \frac{\gamma}{2}\Delta\Big(\ln\frac{c_{ij}^m c_{ij}^f}{c_{i0}c_{0j}}\Big) = \frac{\gamma}{2}\Big(\ln\frac{c_{ij}^m\ \bar{c}^m}{\bar{c}_i^m\ \bar{c}^m{}_j} + \ln\frac{c_{ij}^f\ \bar{c}^f}{\bar{c}_i^f\ \bar{c}^f{}_j}\Big) \tag{25}$$

which means that the transferable component of AM is the average of assortative matching in private consumption expenditure of men and women.

This decomposition of AM is a crucial element of our counterfactual analysis. In the following section, we demonstrate that the two components have opposing effects on income inequality. Specifically, the prevailing intuition in the literature, which attributes an inequality-increasing effect to AM, is primarily a feature of its transferable component, not its non-transferable counterpart. Furthermore, analogous to the decomposition of the population table into AM and population vectors, one can derive the income and sharing rule tables by combining the components of AM with the pecuniary and non-pecuniary RE vectors.

5.3 A marriage market with two education levels

To get more insight about how each component of AM and RE affect inequality, we explore a simple matching table with two education levels in which AM can be characterized by a single log odds ratio.

For simplicity, we assume $\gamma_{ij} = 1$. Then, in a 2×2 table for married couples, we obtain

$$\rho_{11} = \rho_{22} = -\rho_{12} = -\rho_{21} = \frac{1}{4} \ln \frac{N_{11} N_{22}}{N_{12} N_{21}} \tag{26}$$

$$\rho^a_{11} = \rho^a_{22} = -\rho^a_{12} = -\rho^a_{21} = \frac{1}{4}(a_{11} + a_{22} - a_{12} - a_{21}) \tag{27}$$

$$\rho^b_{11} = \rho^b_{22} = -\rho^b_{12} = -\rho^b_{21} = \frac{1}{4}(b_{11} + b_{22} - b_{12} - b_{21}) \tag{28}$$

$$\rho^Y_{11} = \rho^Y_{22} = -\rho^Y_{12} = -\rho^Y_{21} = \frac{1}{4} \ln \frac{Y_{11} Y_{22}}{Y_{12} Y_{21}} \sqrt{\frac{\lambda_{11}\lambda_{22}}{\lambda_{12}\lambda_{21}} \frac{(1-\lambda_{11})(1-\lambda_{22})}{(1-\lambda_{12})(1-\lambda_{21})}} \tag{29}$$

We characterize equilibrium given the population marginals, marriage rates and the components of RE and AM. From proposition 1, we can compute population table using $N_{i\cdot}$, $N_{\cdot j}$, μ_i, ω_j, and $\rho_{ij} = \frac{1}{2}(\rho^a_{ij} + \rho^b_{ij}) + \rho^Y_{ij}$. Furthermore, we have $D_{i\cdot} = \varphi^m_i N_{i\cdot}$, $D_{\cdot j} = \varphi^f_j N_{\cdot j}$, $C_{i\cdot} = \phi^m_i N_{i\cdot}$, and $C_{\cdot j} = \phi^f_j N_{\cdot j}$. Then, the equilibrium income and sharing rule tables can be found by solving the system of equations including (26) to (29) and the below equations

$$N_{10} + N_{11} \exp(a_{11}) + N_{12} \exp(a_{12}) = D_{1\cdot}, \qquad N_{01} + N_{11} \exp(b_{11}) + N_{21} \exp(b_{21}) = D_{\cdot 1}$$

$$N_{20} + N_{21} \exp(a_{21}) + N_{22} \exp(a_{22}) = D_{2\cdot}, \qquad N_{02} + N_{12} \exp(b_{12}) + N_{22} \exp(b_{22}) = D_{\cdot 2}$$

$$N_{10} Y_{10} + N_{11}\lambda_{11} Y_{11} + N_{12}\lambda_{12} Y_{12} = C_{1\cdot}, \qquad N_{01} Y_{01} + N_{11}(1-\lambda_{11}) Y_{11} + N_{21}(1-\lambda_{21}) Y_{21} = C_{\cdot 1}$$

$$N_{20} Y_{20} + N_{21}\lambda_{12} Y_{12} + N_{22}\lambda_{22} Y_{22} = C_{2\cdot}, \qquad N_{02} Y_{02} + N_{12}(1-\lambda_{12}) Y_{12} + N_{22}(1-\lambda_{22}) Y_{22} = C_{\cdot 2}$$

$$\rho_{ij} = \frac{1}{2}(\rho^a_{ij} + \rho^b_{ij}) + \rho^Y_{ij} \qquad \lambda_{ij} = \frac{Y_{i0} N_{0j} e^{b_{ij}}}{Y_{i0} N_{0j} e^{b_{ij}} + Y_{0j} N_{i0} e^{a_{ij}}} \qquad Y_{ij} = N_{ij}\Big(\frac{Y_{i0}}{N_{i0}} e^{a_{ij}} + \frac{Y_{0j}}{N_{0j}} e^{b_{ij}}\Big)$$

The above system does not always have a solution. For example, since $N_{10} + N_{11} + N_{12} = N_{1\cdot}$, if $D_{1\cdot} < N_{1\cdot} - N_{10}$, no real solution for a_{11} and a_{12} exist. When the inputs of the model are such that we have an equilibrium, the average consumption expenditure by individual is

$$\bar{Y} = \frac{C_{1\cdot} + C_{\cdot 1} + C_{2\cdot} + C_{\cdot 2}}{N_{1\cdot} + N_{\cdot 1} + N_{2\cdot} + N_{\cdot 2}} \tag{30}$$

and we measure ***inequality at individual level*** by total required transfer to reach perfect equality for individual consumption

$$\begin{aligned} T = & \sum_{i=1}^{2} N_{i0} \mathbb{1}\big(Y_{i0} > \bar{Y}\big)\big(Y_{i0} - \bar{Y}\big) + \sum_{j=1}^{2} N_{0j} \mathbb{1}\big(Y_{0j} > \bar{Y}\big)\big(Y_{0j} - \bar{Y}\big) \\ & + \sum_{i=1}^{2}\sum_{j=1}^{2} N_{ij}\Big(\mathbb{1}\big(\lambda_{ij} Y_{ij} > \bar{Y}\big)\big(\lambda_{ij} Y_{ij} - \bar{Y}\big) + \mathbb{1}\big((1-\lambda_{ij}) Y_{ij} > \bar{Y}\big)\big((1-\lambda_{ij}) Y_{ij} - \bar{Y}\big)\Big) \end{aligned} \tag{31}$$

We can also compute ***inequality at household level*** for married couples as

$$\bar{Y}_h = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} N_{ij} Y_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J} N_{ij}} \qquad T = \sum_{i=1}^{2}\sum_{j=1}^{2} N_{ij} \mathbb{1}\big(Y_{ij} > \bar{Y}_h\big)\big(Y_{ij} - \bar{Y}_h\big) \tag{32}$$

Since a general analytical solution to the above system is not available, we use simulations to assess the impact of various factors on income inequality. Appendix A.2 provides a detailed explanation of the solution process and the simplifications applied when the market is symmetric. Here, we summarize how income inequality responds to changes in the components of AM and RE.

Figure 2 illustrates the effect of the two components of AM on inequality, measured either by consumption expenditure or married household income, under different levels of pecuniary RE. For clarity, the inequality index is normalized to its value at the maximum level of the x-axis variable, ensuring all indices converge to one at the terminal value. In these simulations, we assume constant population, marriage rates, and non-pecuniary RE, along with symmetric pecuniary RE values, calibrated to align with real-world data.

We illustrate the curves for the AM components changing between 0 (random matching) and 3, which is the max of diagonal elements of AM in the US data. The left-side graphs show that an increase in the non-transferable component of AM consistently decreases inequality, regardless of whether it is measured at the household or individual level. In contrast, the right-side graphs illustrate that the relationship between the transferable component of AM and inequality is generally non-monotonic. For small values of ρ^Y, inequality indices initially rise but for bigger values, the relationship reverses after a certain threshold at lower levels of pecuniary RE and inequality begins to decline. Thus, while both inequality measures always rise when ρ^Y is around zero, their eventual behavior depends on the level of pecuniary RE.

Figure 3 illustrates the effect of changes in the relative non-pecuniary return to education ($\frac{\varphi_2}{\varphi_1}$) on inequality across various levels of the average non-pecuniary RE index. The simulations assume constant population, marriage rates, and symmetric pecuniary RE. The simulate the relationship for values of $\frac{\varphi_2}{\varphi_1}$ range between $[0, 2]$, but as the figure shows, the system has no real solution for a significant portion of values below 1. As the average non-pecuniary RE index increases, the system converge to a solution over a wider domain below 1.

The trends reveal that inequality reaches its minimum when low- and high-education groups yield equal non-pecuniary RE. Specifically, inequality decreases as the ratio of non-pecuniary RE shifts in favor of the low-educated group but increases when it favors the high-educated group. This indicates that any imbalance in non-pecuniary RE between the two education groups, regardless of direction, contributes to greater income inequality at equilibrium.

Figure 4 illustrates the impact of the pecuniary component of RE on inequality. The simulations assume consistent population, marriage rates, and non-pecuniary RE, along with symmetric pecuniary RE, calibrated to reflect real data. Unlike the non-pecuniary return, the average pecuniary return to education does not influence income inequality and the graphs remain unchanged for different values of $\bar{\phi}$.

Since higher education generally yields greater pecuniary returns—a pattern supported by real data—we

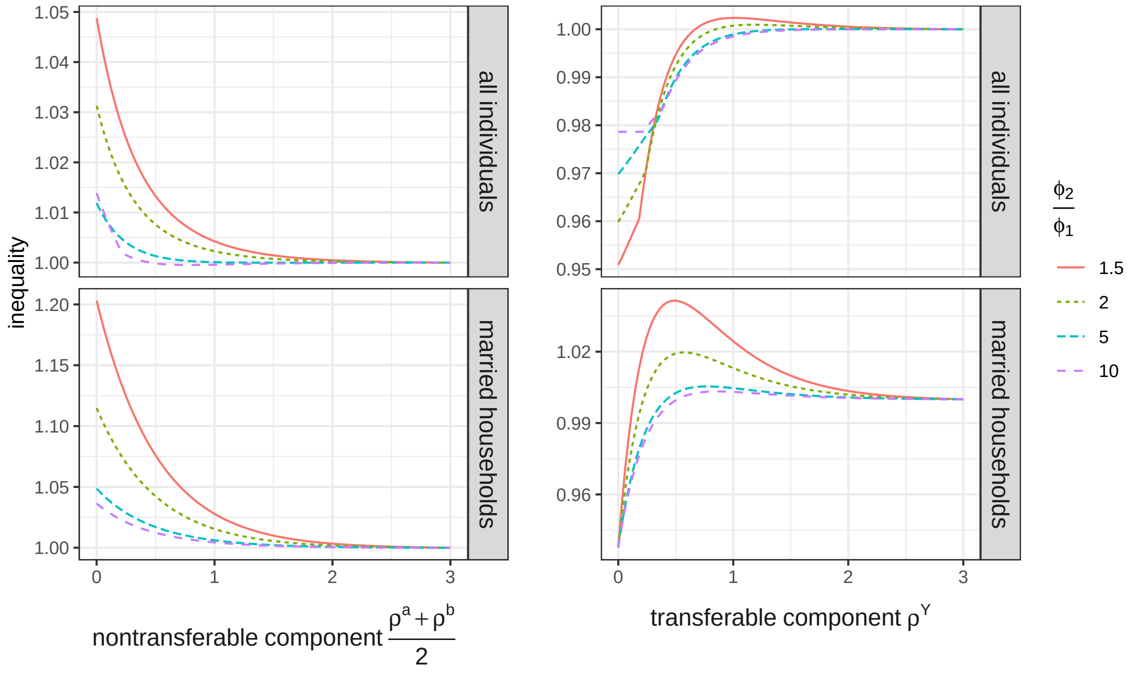

Figure 2: The relationship between two components of AM and inequality. Inequality is measured by total required transfer to reach perfect equality and is normalized to its value at the max level of AM variable. The inequality for all individuals is computed based on their share of total income as in (31) and for married households based on household income as in (32). The calibrated parameters are $D_i = D_{j} = N_i = N_{j} = 8$, $\mu_i = \omega_j = 0.7$, $C_i = C_{i}$ and $C_1 + C_2 = 1000$. For illustrating the trend of non-transferable AM, the value of transferable component is considered 1 and vice-versa.

focus on simulations for $\frac{\phi_2}{\phi_1} \geq 1$.[7] The results suggest a positive relationship between the ratio of pecuniary RE and income inequality. This finding aligns with the non-pecuniary RE pattern when it favors more educated individuals. However, a key distinction observed in the data is that while the pecuniary return consistently favors more educated individuals ($\phi_2 > \phi_1$), the non-pecuniary return may favor either education group. In summary, the findings of this section highlight the importance of distinguishing between the two components of AM when analyzing its impact of on income inequality. Specifically, the non-transferable component of AM monotonically decreases inequality, while the transferable component increases it. Furthermore, the effect of changes in relative non-pecuniary RE on inequality depends on its initial level—it can either increase or decrease inequality. In contrast, the pecuniary RE, which increases with education, has a monotonic and always increasing effect on inequality.

5.4 Finding equilibrium in general setting

As discussed in Section 3.3, constructing counterfactual measures of inequality requires estimating population and income tables under various scenarios. To do this, it is essential to establish the timing of the model's components. Since our focus is on decomposing income inequality through the evolution of AM, we must determine how equilibrium outcomes respond to changes in AM. As outlined in Section 5.2, AM

[7] For 0 $\frac{\phi_2}{\phi_1}$ 1, simulations yield results similar to Figure 3, with minimum inequality occurring at $\frac{\phi_2}{\phi_1} = 1$.

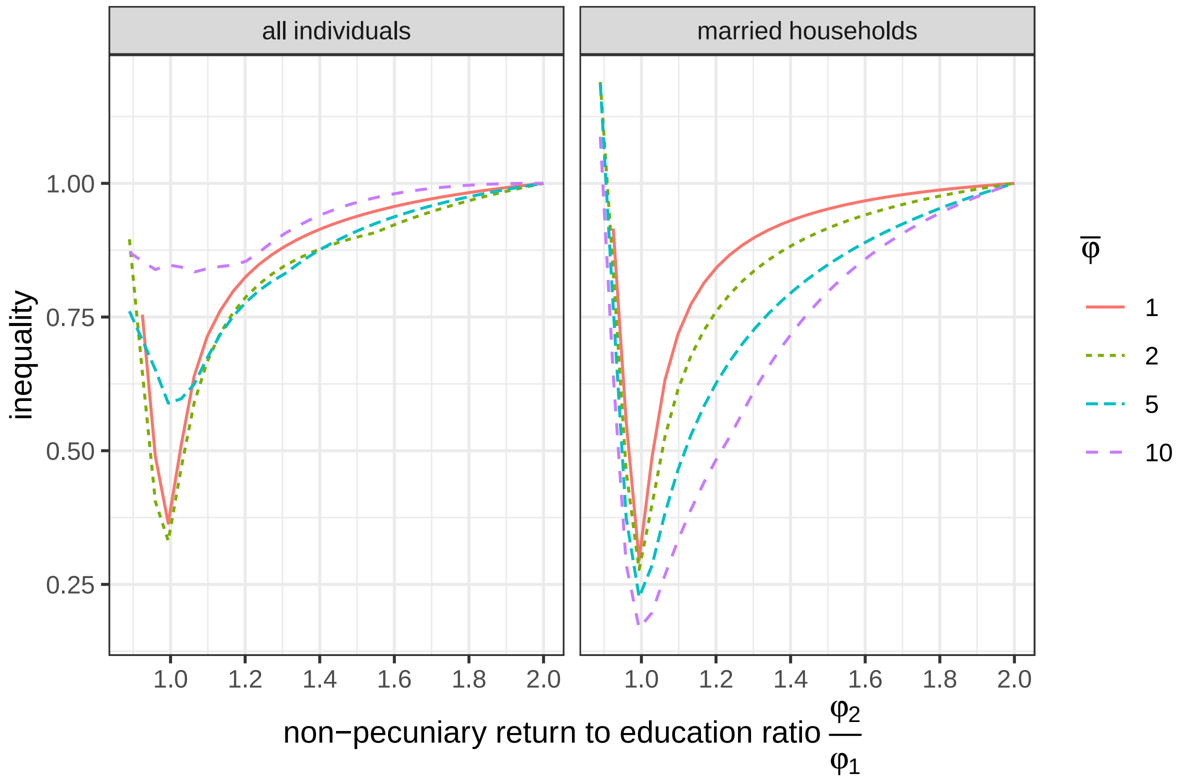

Figure 3: The relationship between the non-pecuniary RE and inequality. Inequality is measured by total required transfer to reach perfect equality and is normalized to its value at the max level of the return variable. Here, we assume $\rho^a_{11} = \rho^b_{11} = 1, \rho^Y_{11} = 0, C_2 \ /C_1 \ = C \ _2/C \ _1 = 2$ and $\mu_i = \omega_j = 0.7$. Different curves correspond to different levels of average non-pecuniary RE $\bar{\varphi} = \frac{D_{1\oplus}+D_{2\oplus}}{N_{1\oplus}+N_{2\oplus}}$. Inequality for all individuals is computed based on their share of total income as in (31) and for married households based on household income as in (32).

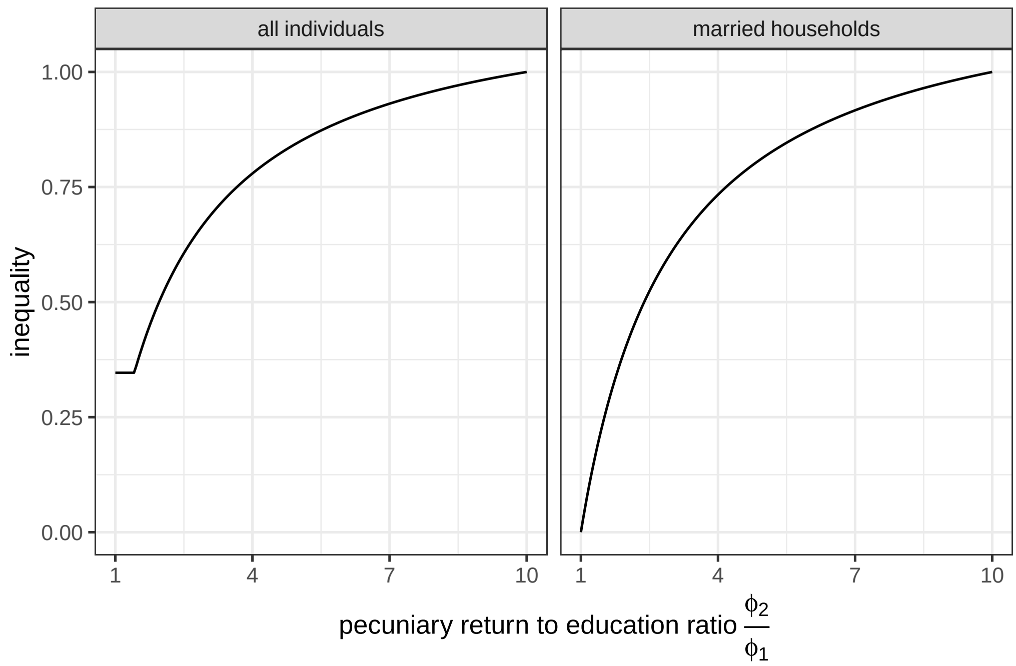

Figure 4: The relationship between pecuniary RE and inequality. Inequality is measured by total required transfer to reach perfect equality and is normalized to its value at the max level of the RE variable. Here, we assume $\rho^a_{11} = \rho^b_{11} = 1, \rho^Y_{11} = 0$ and $\mu_i = \omega_j = 0.7$. The inequality index for all individuals is computed based on their share of total income as in (31) and for married households based on household income as in (32). In contrast to Figure 3, the average of pecuniary RE does not affect income inequality and the graphs for different $\bar{\phi}$ are the same.

comprises two elements: a non-transferable component, $\frac{1}{2}(\rho^a_{ij} + \rho^b_{ij})$, and a transferable component, ρ^Y_{ij}. To replicate the AM of a benchmark year, all elements of these two components must match their values from that year.

Other inputs of the model to find equilibrium are $\boldsymbol{N}$, $\boldsymbol{N}$, $\boldsymbol{\mu}, \boldsymbol{\omega}$, and either $\boldsymbol{\phi}^m, \boldsymbol{\phi}^f, \boldsymbol{\varphi}^m, \boldsymbol{\varphi}^f$ or $\boldsymbol{C}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{D}$. Then, from Proposition 1, we can build population table using AM components $\boldsymbol{\rho}^a, \boldsymbol{\rho}^b, \boldsymbol{\rho}^Y$, marriage rates $(\boldsymbol{\mu}, \boldsymbol{\omega})$, and marginal populations ($\boldsymbol{N}$, $\boldsymbol{N}$). To find income table, let $\bar{x}_i = \frac{1}{J} x_{i+}$, $\bar{x}_j = \frac{1}{I} x_{+j}$, $\bar{x} = \frac{1}{IJ} x_{++}$, and

$$Z_{ij} = \ln \left(Y_{ij} \sqrt{\frac{\lambda_{ij}(1-\lambda_{ij})}{Y_{i0} Y_{0j}}} \right)^{\gamma_{ij}}$$

Then, the following relationships must hold at the equilibrium

$$N_{i0} + \sum_{j=1}^{J} N_{ij} \exp\left(\frac{a_{ij}}{\gamma_{ij}}\right) = D_i \;, \qquad N_{0j} + \sum_{i=1}^{I} N_{ij} \exp\left(\frac{b_{ij}}{\gamma_{ij}}\right) = D_j \tag{33}$$

$$N_{i0} Y_{i0} + \sum_{j=1}^{J} N_{ij} \lambda_{ij} Y_{ij} = C_i \;, \qquad N_{0j} Y_{0j} + \sum_{i=1}^{I} N_{ij} (1-\lambda_{ij}) Y_{ij} = C_j \tag{34}$$

$$\bar{S}_i = \frac{1}{2}(\bar{a}_i + \bar{b}_i) + \bar{Z}_i \;, \qquad \bar{S}_j = \frac{1}{2}(\bar{a}_j + \bar{b}_j) + \bar{Z}_j \tag{35}$$

It is important to note that a solution to the equilibrium of system does not always exist, meaning that some counterfactual scenarios, where one of the model's inputs is replaced with its benchmark year value, may be unattainable.

When a solution does exist, equilibrium can be computed using an optimization algorithm similar to the 2×2 case detailed in Appendix A.2. This process involves expressing income and sharing rules as functions of $\bar{a}_i, \bar{a}_j, \bar{b}_i, \bar{b}_j$ and then minimizing the sum of squared residuals from equations as (33) and (35). However, because the derivative of the objective function cannot be derived analytically, in models with multiple categories, selecting appropriate initial values to ensure convergence becomes a significant computational challenge. To address this, we propose an alternative algorithm described in Appendix A.3, which offers a more efficient approach for initializing and solving the system. In short, we first compute two pairs of functions from the above system

1. $\bar{a}_i(\bar{a}_1, \dots, \bar{a}_{J\text{-}1}, \bar{b}_1, \dots, \bar{b}_{I\text{-}1})$ and $\bar{b}_j(\bar{a}_1, \dots, \bar{a}_{J\text{-}1}, \bar{b}_1, \dots, \bar{b}_{I\text{-}1})$

2. $\bar{a}_j(\bar{a}_1, \dots, \bar{a}_{I\text{-}1}, \bar{b}_1, \dots, \bar{b}_{I\text{-}1})$ and $\bar{b}_i(\bar{a}_1, \dots, \bar{a}_{I\text{-}1}, \bar{b}_1, \dots, \bar{b}_{I\text{-}1})$

Then, we follow the below iterative procedure to find equilibrium $\boldsymbol{a}, \boldsymbol{b}$ and using them compute income and sharing rule tables.

- start with vectors $(\bar{a}_1^{(0)}, \dots, \bar{a}_{J\text{-}1}^{(0)})$ and $(\bar{b}_1^{(0)}, \dots, \bar{b}_{I\text{-}1}^{(0)})$, then at iteration $k \geq 1$
- compute $\bar{a}_i^{(2k\text{-}1)}\left(\bar{a}_1^{(2k\text{-}2)}, \dots, \bar{a}_{J\text{-}1}^{(2k\text{-}2)}, \bar{b}_1^{(2k\text{-}2)}, \dots, \bar{b}_{I\text{-}1}^{(2k\text{-}2)}\right)$ and $\bar{b}_j^{(2k\text{-}1)}\left(\bar{a}_1^{(2k\text{-}2)}, \dots, \bar{a}_{J\text{-}1}^{(2k\text{-}2)}, \bar{b}_1^{(2k\text{-}2)}, \dots, \bar{b}_{I\text{-}1}^{(2k\text{-}2)}\right)$
- compute $\bar{a}_j^{(2k)}\left(\bar{a}_1^{(2k\text{-}1)}, \dots, \bar{a}_{I\text{-}1}^{(2k\text{-}1)}, \bar{b}_j^{(2k\text{-}1)}, \dots, \bar{b}_{J\text{-}1}^{(2k\text{-}1)}\right)$ and $\bar{b}_i^{(2k)}\left(\bar{a}_1^{(2k\text{-}1)}, \dots, \bar{a}_{I\text{-}1}^{(2k\text{-}1)}, \bar{b}_j^{(2k\text{-}1)}, \dots, \bar{b}_{J\text{-}1}^{(2k\text{-}1)}\right)$

The advantage of this algorithm to direct optimization is that the derivative of the above function can be computed analytically which leads to faster convergence.

In summary, building a counterfactual experiment using the above model requires the choice for the timing of the following factors

- t_X: non-transferable AM $(\boldsymbol{\rho}^a, \boldsymbol{\rho}^b)$
- t_Y: transferable AM $(\boldsymbol{\rho}^Y)$
- t_φ: non-pecuniary RE $(\boldsymbol{\varphi}^m, \boldsymbol{\varphi}^f)$
- t_ϕ: pecuniary RE $(\boldsymbol{\phi}^m, \boldsymbol{\phi}^f)$
- t_M: marriage rates $(\boldsymbol{\mu}, \boldsymbol{\omega})$
- t_N: marginal population vectors $(\boldsymbol{N}\ \ , \boldsymbol{N}\ \ )$

However, it is possible that the system does not have a solution for a choice of inputs.

6 Data, Overall Trends, and Parameter Estimations

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,[8] 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 5 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 average income of all individuals in their respective groups. Figure 6 illustrates the trend of average income by gender and education in 1999 dollars by considering zero income for non-participants. The dashed lines 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.

[8]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.

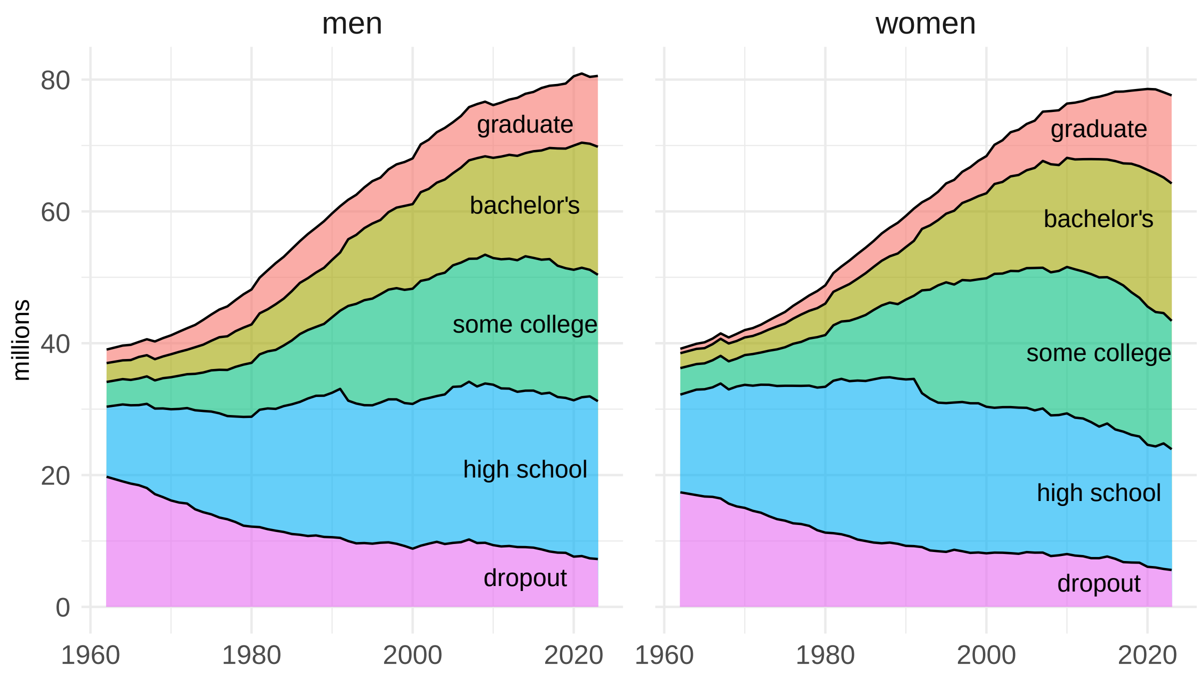

Figure 5: Trend of total population by education and gender in US. Data source: CPS, individuals between 26 and 60 years old.

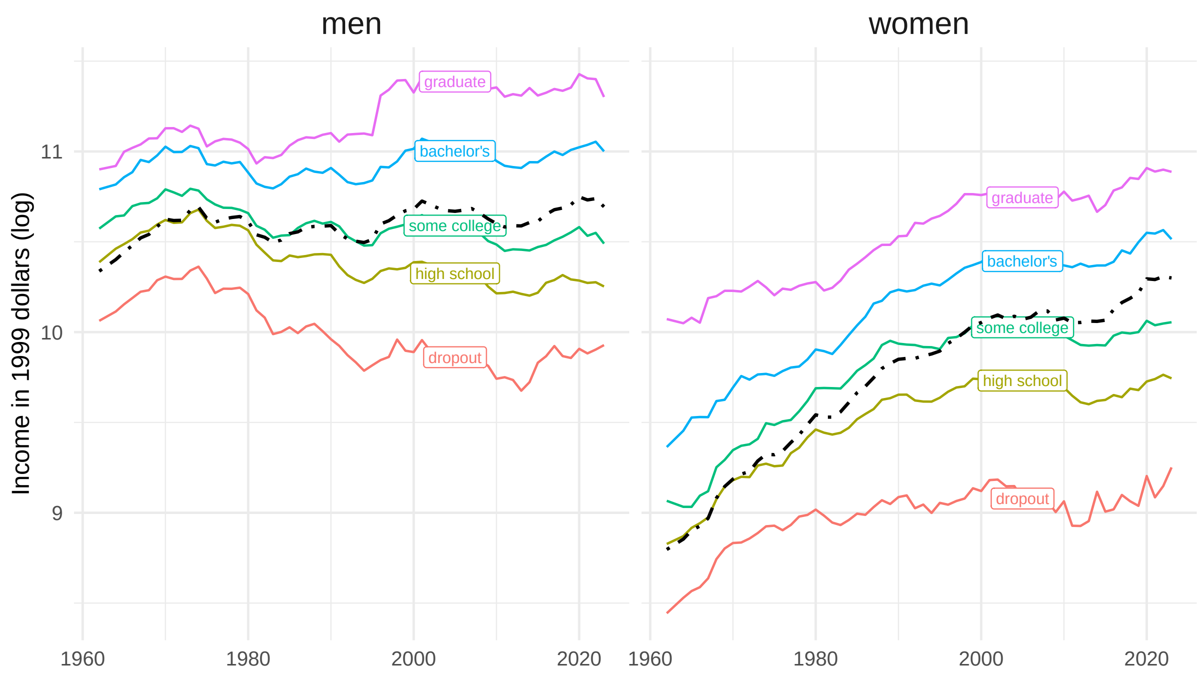

Figure 6: Trend of log average income by education for men and women in the US. For taking average, the income of non-participants are considered as zero. The dashed line show the average across the whole population. Income is adjusted by the 1999 price index. Data source: CPS, individuals between 26 and 60 years old.

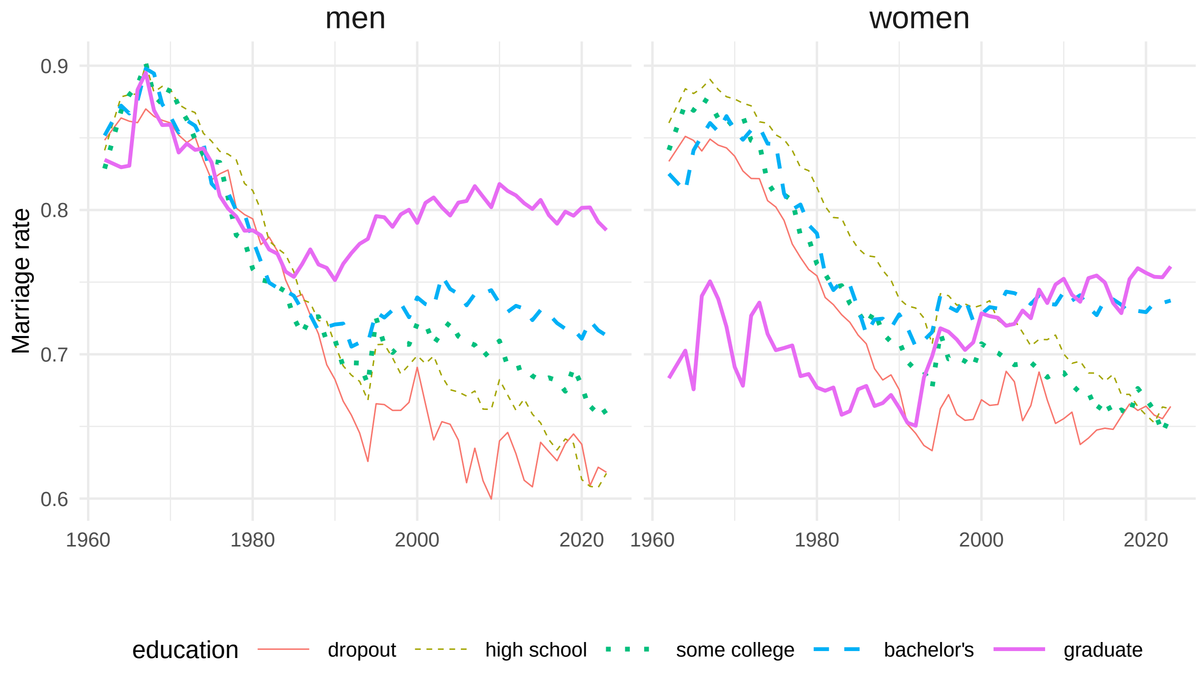

Figure 7: Trend of marriage rate by education for men and women in the US. Data source: CPS, individuals between 26 and 60 years old.

Figure 7 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.

Figure 8 illustrates the trend of AM pattern, defined as ρ_{ij} in (1), 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 18 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.

Figure 9 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.[9] Hence, we can conclude an increasing trend of AM in the US marriage market in the period of study.

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

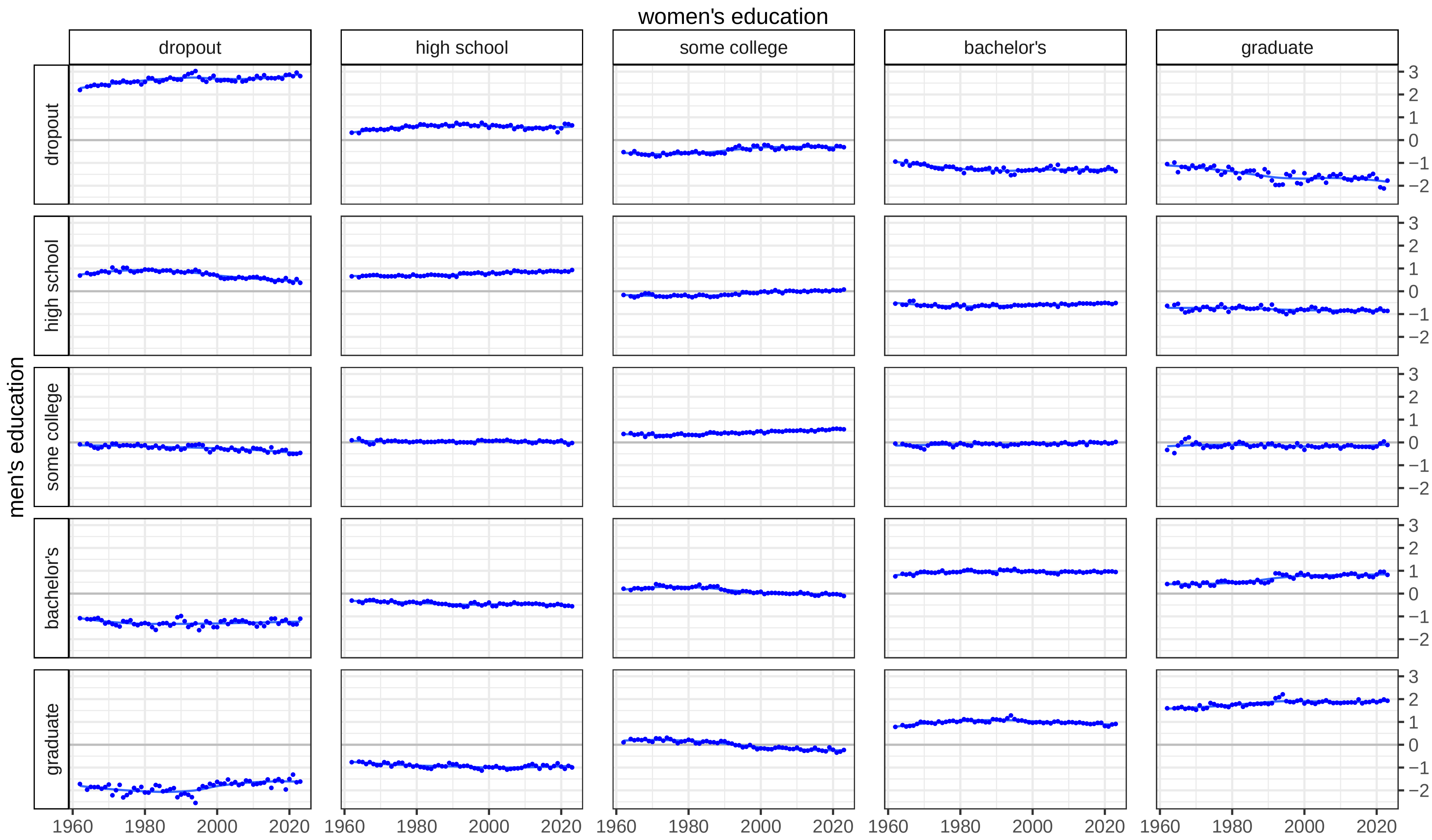

Figure 8: Assortative matching in population measured by log of geometric mean of odds ratios as $\ln \frac{N_{ij} N_{\times\times}}{N_{i\times} N_{\times j}}$.

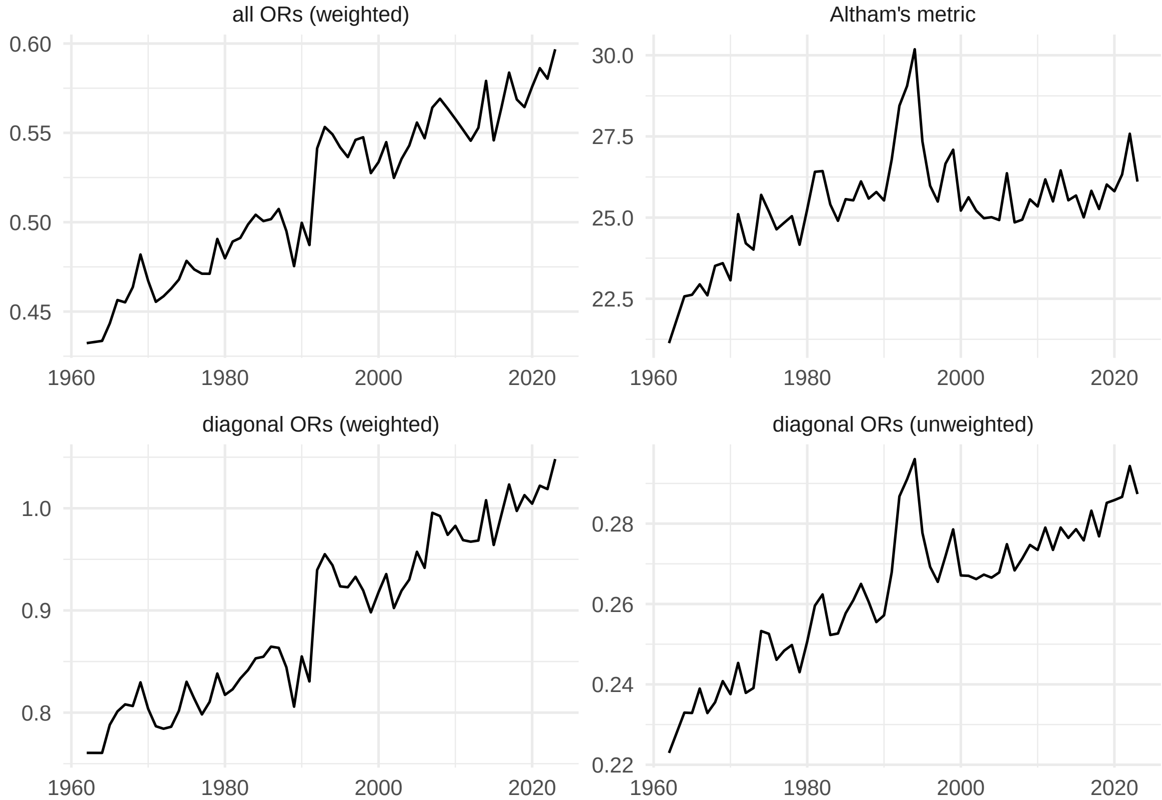

Figure 9: Different aggregate measures of assortativeness using log odds ratios.

6.1 Parameter estimations

As discussed above, γ_{ij} and λ_{ij} are the two structural parameters of the model that cannot be point identified without data on transfer within household and we calibrate them as described in section 4.5.

To estimate γ_{ij} from (19), since couple types with low sample sizes exhibit fluctuations in average income and its variance over time (Appendix Figure 19), we use the smoothed version of $\text{Var}(\ln Y_{ij})$ through non-parametric LOESS regression. The top plot of Figure 10, 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. In the following, in addition the time varying values of γ_{ij}, we repeat the counterfactual estimations using alternative values of γ_{ij} for each couple overtime.

We illustrate the trend of λ_{ij} in bottom plot of Figure 10. 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 estimations, to identify non-pecuniary gains, we use $\lambda_{ij} = \lambda_{ij}$ and as robustness test, we repeat estimation using the average of λ_{ij} and $\frac{1}{2}$ as the sharing rule.

Figure 11, illustrates the estimation of the non-pecuniary parameters a_{ij} and b_{ij} using (15) and compares them with the average surplus $S_{ij} = \ln N_{ij} - \frac{1}{2}\ln(N_{i0}N_{0j})$. We observe that the estimated

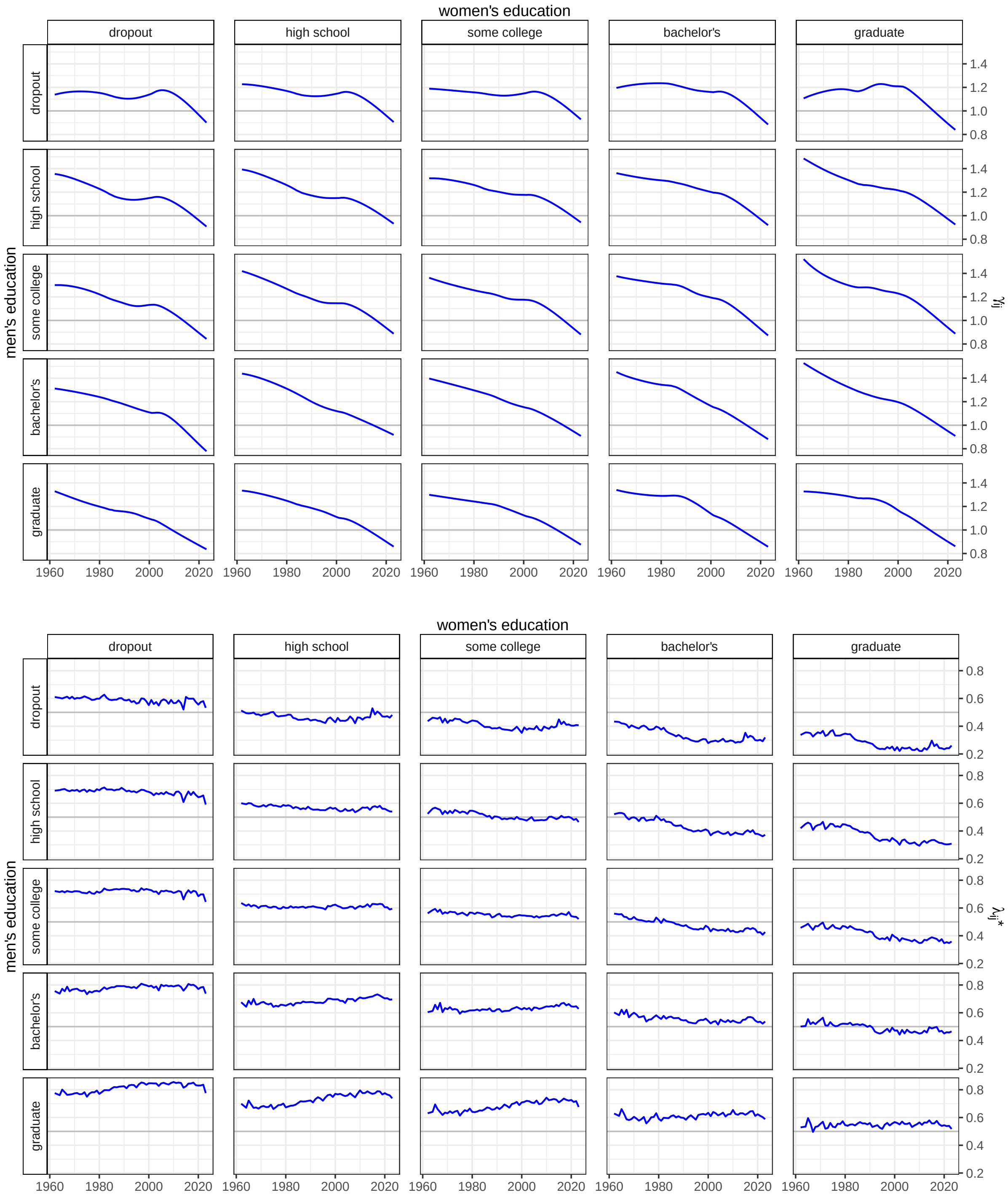

Figure 10: The estimated levels of γ_{ij} and λ_{ij}.

values of non-pecuniary 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 higher-educated partner receives more non-pecuniary gain compared to the lower-educated partner.

Figure 12 shows the trend of the two RE indices estimated as (21) and (23). The top graphs show that non-pecuniary RE has been decreasing overtime in almost all groups. Moreover, in the 1960s, non-pecuniary RE is decreasing in education level with a large gap between different groups, but over time the gap disappears. In the recent year, the non-pecuniary return of graduate education has the highest value for men and the second highest for women. In the bottom graphs, we observe a close relationship between the pecuniary RE index and average income for men (Figure 6), but for women we observe that the lower average income is largely compensated by higher expenditure share. In addition, the trends suggests that the pecuniary RE index is increasing in education is all years and all education levels.

Figure 13 shows the trend of non-transferable and transferable components of AM overtime aggregated as (3). It can clearly seen that the main driver of AM and its change over time is its non-transferable component and the transferable component revolves around zero. Between 1962 and 2023, both aggregate measures suggest about 25% increase in non-transferable AM. The aggregate transferable component, starts from zero in 1962, then becomes slightly negatively assortative in 1970s, and afterwards fluctuates near zero. In the Appendix Figure 20, we separately depict the trend of the non-transferable and transferable components for each element of them and observe that unlike the non-transferable component, there is no clear pattern across diagonal or non-diagonal elements of transferable gains. Thus, these trends highlight an important fact in the US marriage market in the past 60 years: *Assortative matching and its growth overtime has been mainly due to non-transferable gains of marriage and not its transferable gains.*

7 Counterfactual Experiments

In this section, we conduct counterfactual experiments to examine the contributions of two components of AM and RE on income inequality in the US over the period 1962-2003. To measure income inequality, we use the Gini coefficient for these samples:

- sample of married couples,
- sample of all individuals where we divide the household income of married couples of each type between the partners, according to their average sharing rule λ_{ij}.

The estimated trend of Gini coefficient, illustrated by thick lines in the subsequent graphs, suggest that, on average, inequality is higher when measured at the individual level, particularly in the earlier years of

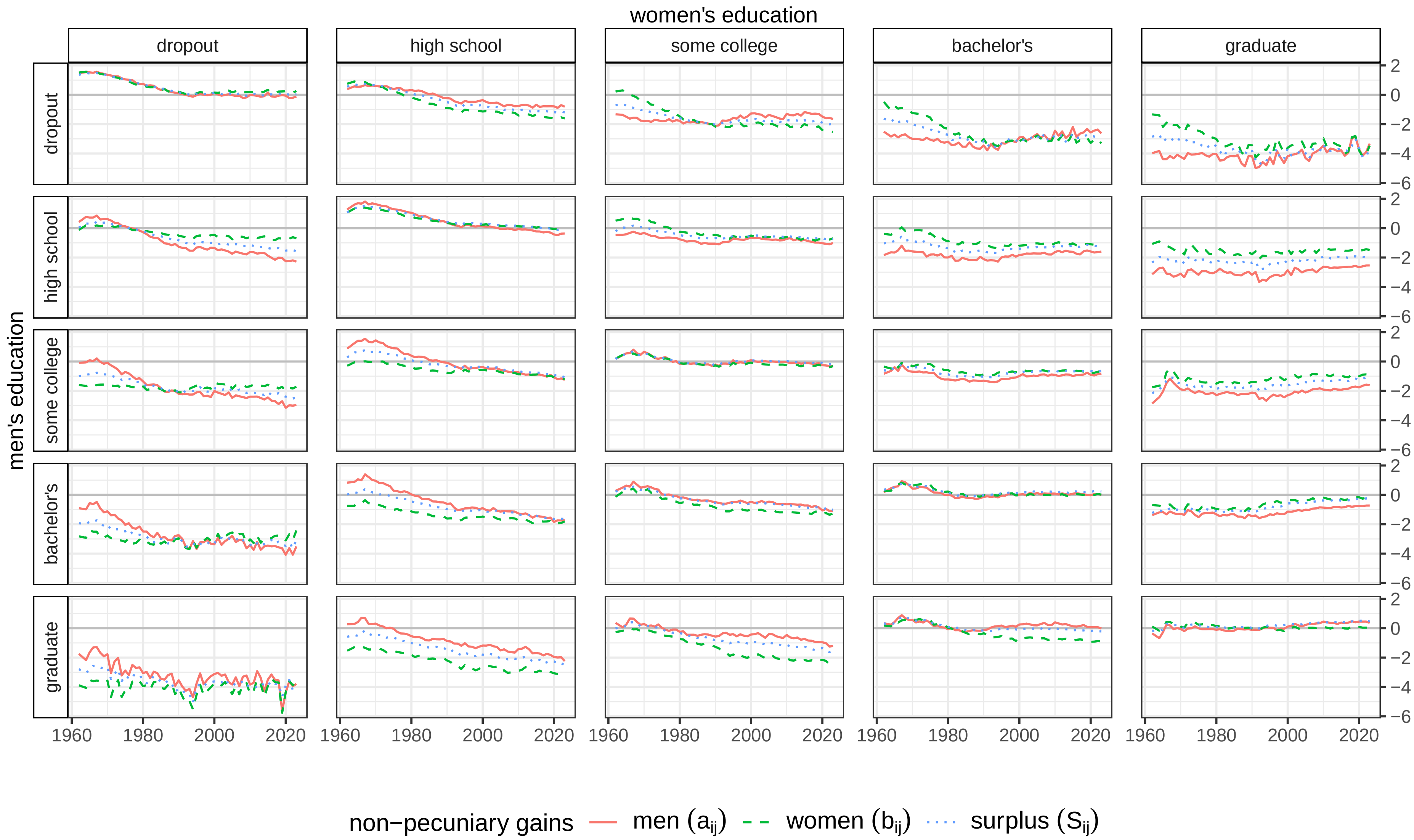

Figure 11: The estimated levels of non-pecuniary gains a_{ij} and b_{ij}, and total surplus S_{ij}.

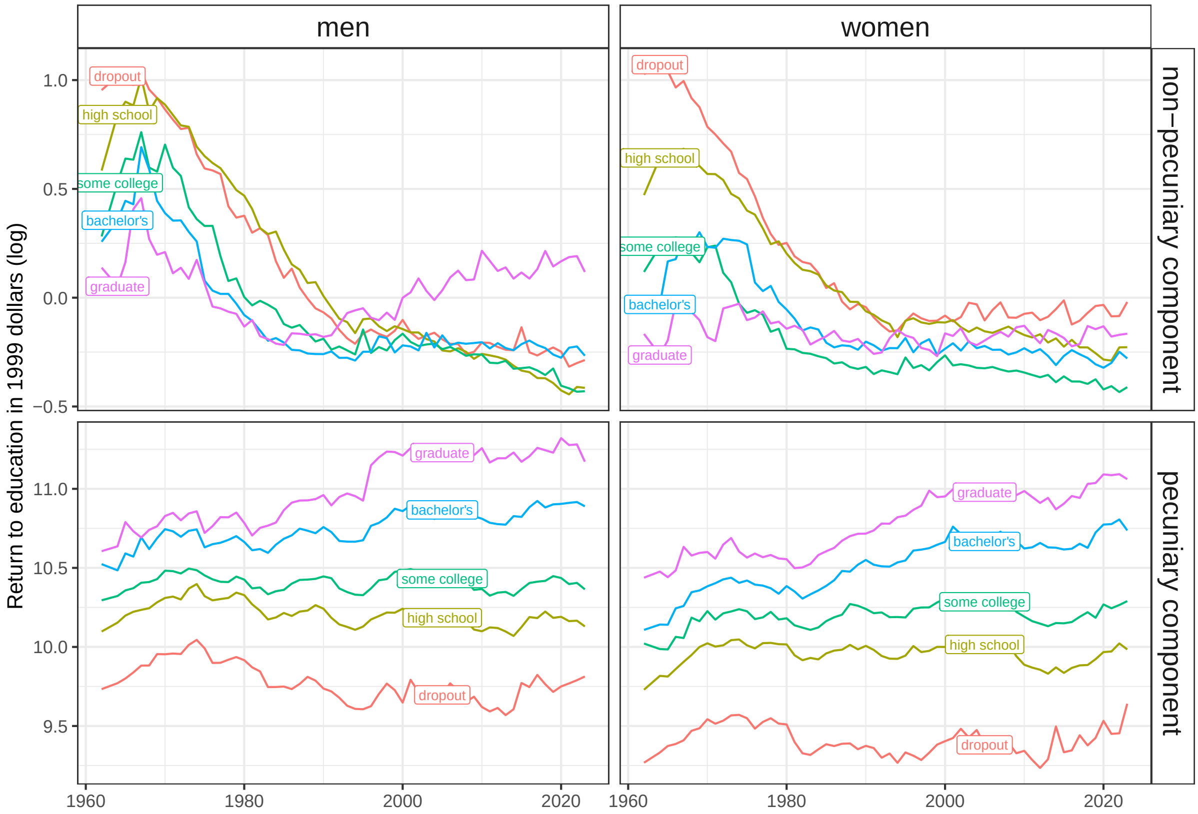

Figure 12: The trends of pecuniary and non-pecuniary components of return to education in the US.

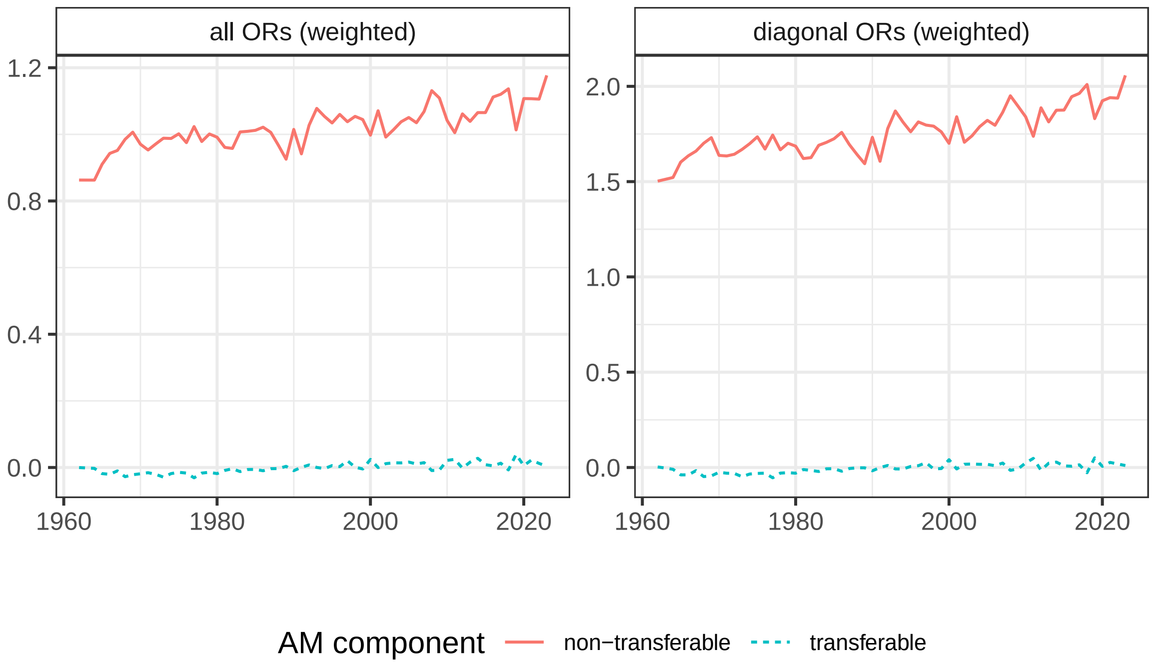

Figure 13: The trends of aggregate non-pecuniary $(\frac{\boldsymbol{\rho}^a+\boldsymbol{\rho}^b}{2})$ and pecuniary $(\boldsymbol{\rho}^Y)$ components of AM as defined in (24) and aggregated as (3).

the data. In particular, the Gini index is 0.328 and 0.365 at 1962 and 0.421 and 0.442 at 2023, for the sample of married couples and individuals, respectively.

7.1 The US income inequality with constant AM

The top and bottom graphs of Figure 14 show the estimated actual and counterfactual inequality for the samples of married households and all individuals under different scenarios for AM. The thick lines shows the actual levels of Gini coefficient, and the others illustrate the outcome inequality in three counterfactual exercises regarding the components of AM for each sample: when either non-transferable or transferable AM and when both of them are at their 1962 levels.

Fixing only the non-transferable component of AM to its 1962 level leads to an increase in inequality, that in 2023 its numbers are approximately 10 and 14 Gini points growth, for the sample of married couples and all individuals, respectively. In contrast, fixing only the transferable component of AM to its 1962 level either does not have a solution or it merely changes income inequality. As the main driver of AM is its non-transferable component (Figure 9), when both components of AM are at their 1962 level, the counterfactual inequality is close to its value when only its non-transferable component is at the benchmark year.

These two counterfactual outcomes are consistent with the relationship of their respective component with inequality as depicted in Figure 2. Figure 9 shows that while AM due to non-transferable gain has significantly increased, transferable AM is nearly zero and its change overtime is insignificant. According to section 5.3, the rise of non-transferable AM is associated with less inequality, so when its pattern for recent years (higher AM) is substituted by 1962 pattern (lower AM), inequality rises. In comparison, while the impact of transferable AM on inequality is mostly increasing, its low level and low variation does not change inequality.

To examine the robustness of these finding, we repeat them with alternative levels of γ_{ij} and λ_{ij} for the estimation of non-transferable gains a_{ij}, b_{ij} and marginals $D_{i\ }, D_{\ j}, C_{i\ }, C_{\ j}$. Figure 17 shows the impact of different levels of transferibility parameters on the finding. Here, we consider the same level of γ_{ij} for all types and years and change this level to observe how the counterfactual inequality with fixed AM of 1962 behave based on transerability of the utilities between households. We observe that when the model approches to TU ($\gamma_{ij} \to +\infty$) the counterfactual inequality approches to its actual level. In contrast, approaching to an NTU model widens the gap between actual and counterfactual Gini coefficients. This outcome stems from the fact that γ_{ij} is the parameter to determine the weight of the two components of AM. When it approach zero, AM is fully determined by its non-transferable component ($\boldsymbol{\rho} \approx \frac{\boldsymbol{\rho}^a + \boldsymbol{\rho}^b}{2}$) and when it approach $+\infty$, $\boldsymbol{\rho} \approx \boldsymbol{\rho}^Y$. Since changes in AM is due to its non-transferable component and its transferable component remains near zero, by increasing transferibility the share of low-variation component of AM rises and we obtain lower change in the counterfactual inequality. In further sensitivity analysis, Appendix Figure 21, instead of λ_{ij}, uses the average of λ_{ij} and $\frac{1}{2}$ as the

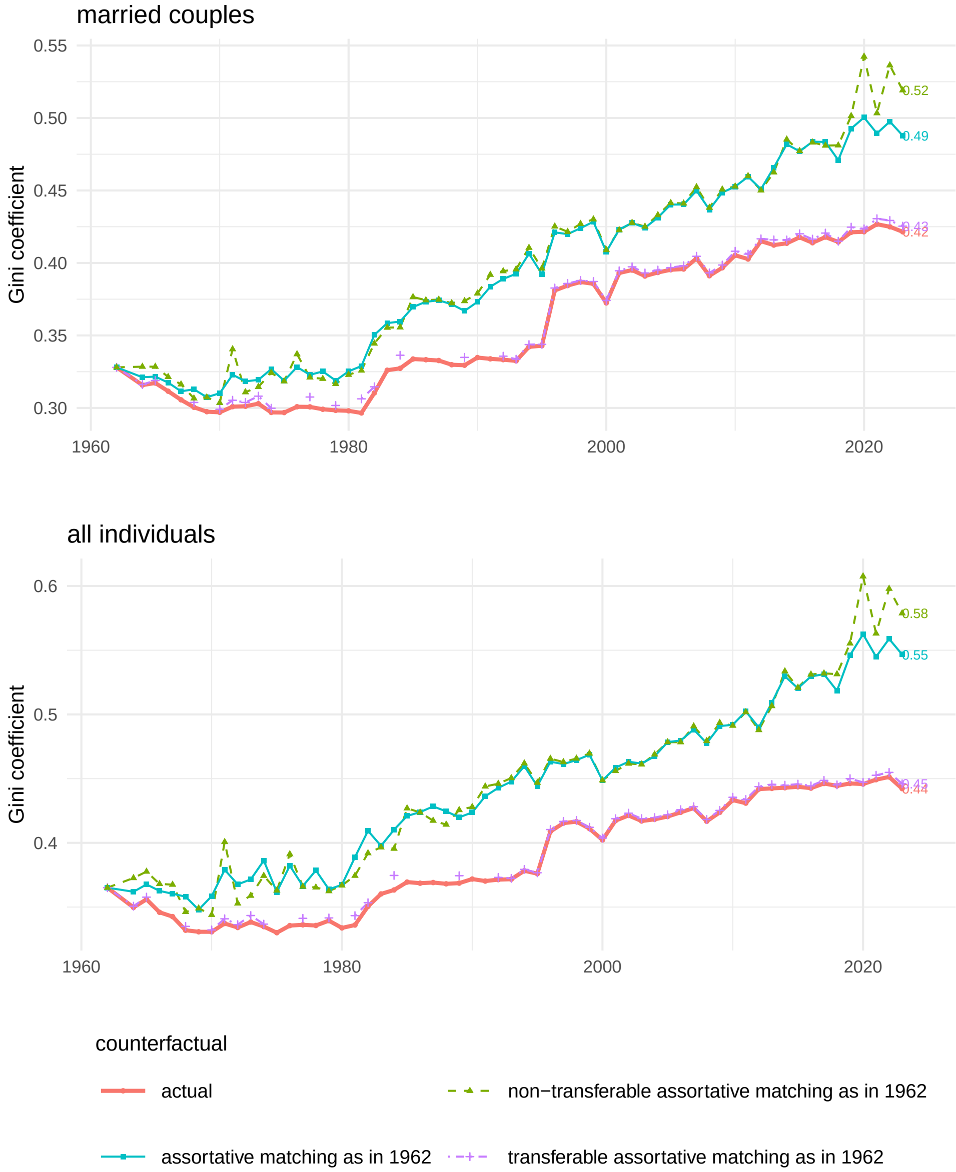

Figure 14: Counter-factual inequality under different scenarios of AM components. The top graph is at household level and in the bottom graph married couples of each type are counted as two individual with income according to the sharing rule λ_{ij}.

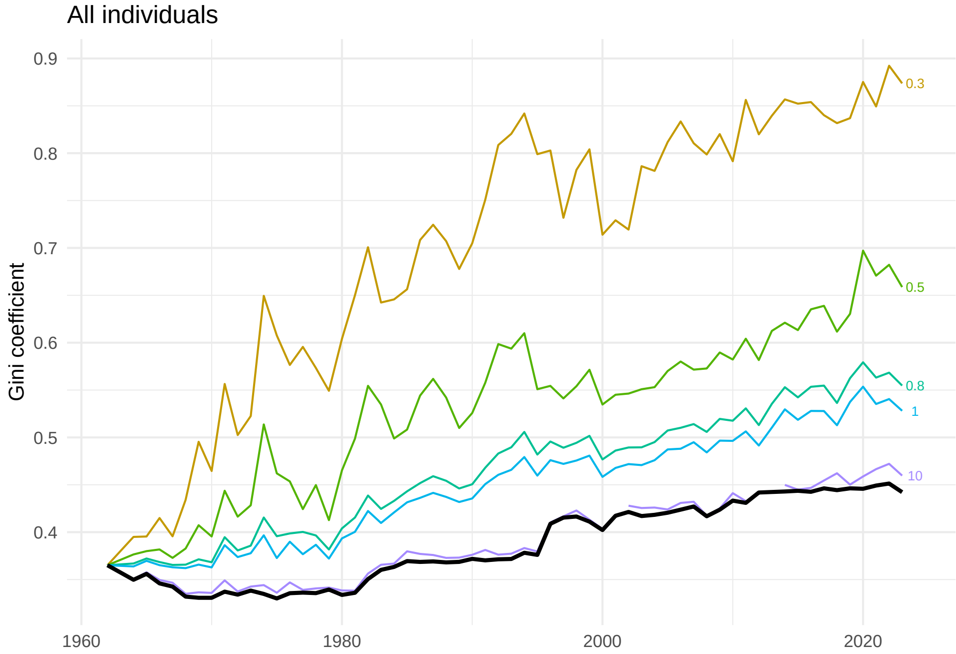

Figure 15: Counterfactual income inequality with AM as in 1962 and different level of transferibility parameter γ_{ij}. The thick line is the actual trend and the others are counterfactual trends. When the model approaches TU (large γ_{ij}) inequality does not change by AM, but with NTU ($\gamma \to 0$) inequality increases more in the counterfactual scenario.

sharing rule, and suggest no significant change in the findings.

7.2 The US income inequality with constant RE

Our second set of exercises examines the contribution of the two RE indices on inequality. Figure 16 shows the outcomes of the three exercises. We observe that fixing non-pecuniary RE at the 1962 levels result in higher income inequality in all following years that a solution exists given the inputs of the system. This finding is consistent with the trends in top graphs of Figure 12 and the relationship depicted in Figure 3. The ranking of the φ_i^m, φ_j^f in 1962 is such that the lower education has higher return and we are at the decreasing segment of Figure 3 when the relative return is smaller than one. In 1970s and 80s, there is a gradual decrease in the return and after 1990 there is no clear ranking by education level, which corresponds to interval near one in Figure 3. Therefore, fixing non-pecuniary RE at the year which is substantially in favor of lower education, leads to higher counterfactual inequality in years that non-pecuniary return is originally equal across education groups.

Regarding the counterfactual exercise in which the pecuniary RE index is at 1962 level, we find a slightly higher Gini coefficient before 1980s, but afterward return to education significantly contributes in explaining the rise in income inequality. Specifically, based on this counterfactual, controlling for pecuniary RE accounts for 34%, 39%, and 40%, of the overall increase in income inequality between 1962 and 2023 for the sample of all household, married couples, and individuals, respectively. Fixing both RE indices at their 1962 level gives the net effect of the two components and leads to higher income inequality, but at lower levels than when only non-pecuniary RE is at benchmark year.

8 Conclusion

This paper investigates the relationship between educational sorting in the marriage market and cross-sectional income inequality, focusing on the role of return to education in shaping both. A frictionless matching model with imperfectly transferable utility is developed to decompose assortative matching into transferable and non-transferable components. While the transferable component, primarily reflecting income-driven sorting preferences, tends to increase inequality, the non-transferable component, encompassing non-pecuniary factors such as affinity and children, mitigates inequality. This result challenges the common belief that assortative matching inevitably exacerbates income inequality.

Using U.S. Current Population Survey (CPS) data from 1962 to 2023, the study empirically demonstrates that the rise in assortative matching primarily stems from its non-transferable component, implying that, after accounting for return to education, marriage market sorting has actually reduced cross-sectional income inequality. The findings also highlight the dominant role of market return to education in driving inequality, accounting for approximately 40% of the increase in the Gini coefficient between 1962 and 2023.

The significant contribution of the non-transferable component of assortative matching suggests that the incentive of highly educated couples to invest in their children's human capital, may inadvertently play a role in reducing cross-sectional income inequality. However, this tendency could potentially intensify long-term inequality as it translates into greater inequality in next generations. These findings underscore the complex relationship between marriage market dynamics and income inequality and ask for further research to investigate the dynamic relationship between education, martial sorting, and income inequality across generations.

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.

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

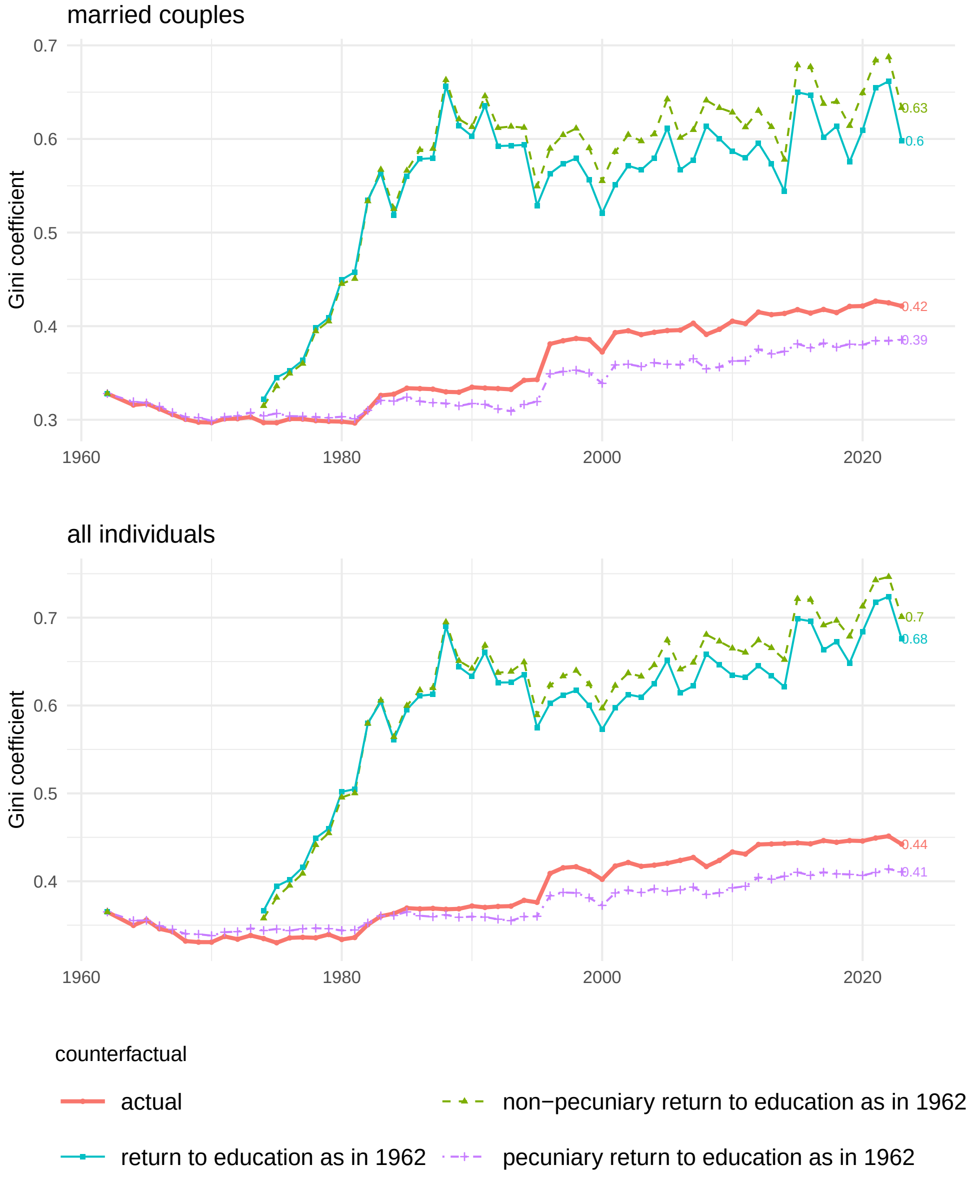

Figure 16: Counter-factual inequality in different scenarios. The top and middle graphs are at household level and in the bottom graph married couple of each type are counted as two individual with income according to the sharing rule λ_{ij}.

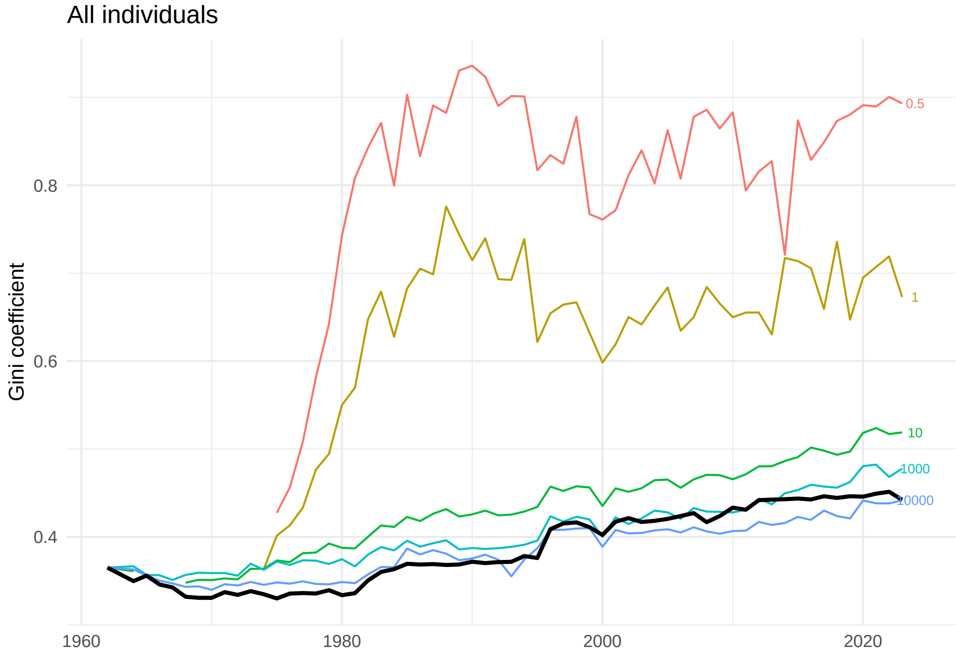

Figure 17: Counterfactual income inequality with FE as in 1962 and different level of transferibility parameter γ_{ij}. When the model approaches TU (large γ_{ij}) inequality approaches to a counterfacutal in which only pecuniary RE is fixed.

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 (2020): "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 (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.

Ciscato, E. and S. Weber (2020): "The role of evolving marital preferences in growing income inequality," *Journal of Population Economics*, 33, 307–347.

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.

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 (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*.

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

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 $N_{i+} = \mu_i N_i\ \ , N_{+j} = \omega_j N\ \ _j$, 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 $m, m^0 \in i$ and $f, f^0 \in j$ such that under stable matching m and m^0 respectively match with f and f^0. 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,

$$D_{ij}(u_m - \alpha^j_m, v_f - \beta^i_f) \leq D_{ij}(u_{m'} - \alpha^j_{m'}, v_f - \beta^i_f) \tag{36}$$
$$D_{ij}(u_{m'} - \alpha^j_{m'}, v_{f'} - \beta^i_{f'}) \leq D_{ij}(u_m - \alpha^j_m, v_{f'} - \beta^i_{f'}) \tag{37}$$

Based on Lemma 1 of GKW, $D_{ij}(u, v)$ is isotone in the sense that $(u, v) \leq (u^0, v^0)$ implies $D_{ij}(u, v) \leq D_{ij}(u^0, v^0)$ and vice-versa. Based on this property of $D_{ij}(\cdot, \cdot)$, from (36), we get $u_m - \alpha^j_m \leq u_{m'} - \alpha^j_{m'}$ and from (37), 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 $D_{ij}(U_{ij}, V_{ij}) = 0$. $\square$

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

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

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

$$\begin{aligned} \Pr\{m \in i,\ j = \arg\max u^k_m\} &= \Pr\{\forall k,\ \alpha^k_m \le U_{ij} - U_{ik} + \alpha^j_m\} \\ &= \int_{-\infty}^{+\infty} \prod_{j \neq k} F_{\alpha^k}(U_{ij} - U_{ik} + \alpha^j_m)\, f_{\alpha^j}(\alpha^j_m)\, d\alpha^j_m = \frac{N_{ij}}{N_i} \qquad (38) \\ \Pr\{f \in j,\ i = \arg\max v^k_f\} &= \Pr\{\forall k,\ \beta^k_f \le V_{ij} - V_{kj} + \beta^i_f\} \\ &= \int_{-\infty}^{+\infty} \prod_{i \neq k} F_{\beta^k}(V_{ij} - V_{kj} + \beta^i_f)\, f_{\beta^i}(\beta^i_f)\, d\beta^i_f = \frac{N_{ij}}{N_{\cdot j}} \qquad (39) \end{aligned}$$

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

$$\frac{\exp(U_{ij})}{\sum_{k=0}^{J} \exp(U_{ik})} = \frac{N_{ij}}{N_i} \qquad \text{and} \qquad \frac{\exp(V_{ij})}{\sum_{k=0}^{I} \exp(V_{kj})} = \frac{N_{ij}}{N_{\cdot j}} \qquad (40)$$

and we then obtain

$$U_{ij} - U_{i0} = \ln \frac{N_{ij}}{N_{i0}}, \qquad V_{ij} - V_{0j} = \ln \frac{N_{ij}}{N_{0j}} \qquad (41)$$

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 (41) leads to

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

$\square$

Proposition 4. From (6)

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

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

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

□

A.2 Finding equilibrium of the 2×2 Model

To solve the system of equations, we define an objective function with arguments $\bar{a}_i$, $\bar{a}$ $_j$, $\bar{b}_i$, $\bar{b}$ $_j$ that if converges to zero the system have a solution. The inputs are $\rho^a_{11}, \rho^b_{11}, \rho_{11}$ (or ρ^Y_{11}), N_i , N $_j$,D_i , D $_j$, C_i , C $_j$, μ_i, ω_j. Since the matrix is 2×2, AM matrix is identified by one element, we drop the sub-indices from the AM elements in the rest of this section. In the first step, using AM indices, marriage rates, and marginal population, we compute population table. Then, we have

$$
a_{11} = \rho^a + \bar{a}_1 \; + \bar{a}\;_1 - \bar{a} \;\;, \quad a_{12} = -\rho^a + \bar{a}_1 \; + \bar{a}\;_2 - \bar{a} \;\;, \quad a_{21} = -\rho^a + \bar{a}_2 \; + \bar{a}\;_1 - \bar{a} \;\;, \quad a_{22} = \rho^a + \bar{a}_2 \; + \bar{a}\;_2 - \bar{a}
$$

$$
b_{11} = \rho^b + \bar{b}_1 \; + \bar{b}\;_1 - \bar{b} \;\;, \quad b_{12} = -\rho^b + \bar{b}_1 \; + \bar{b}\;_2 - \bar{b} \;\;, \quad b_{21} = -\rho^b + \bar{b}_2 \; + \bar{b}\;_1 - \bar{b} \;\;, \quad b_{22} = \rho^b + \bar{b}_2 \; + \bar{b}\;_2 - \bar{b}
$$

$$
Y_{10} = \frac{C_1}{N_{10}\Big(1 + (\frac{N_{11}}{N_{10}})^2 e^{\;\; a_{11}} + (\frac{N_{12}}{N_{10}})^2 e^{\;\; a_{12}}\Big)}, \qquad Y_{01} = \frac{C \;\; _1}{N_{01}\Big(1 + (\frac{N_{11}}{N_{01}})^2 e^{\;\; b_{11}} + (\frac{N_{21}}{N_{01}})^2 e^{\;\; b_{21}}\Big)}
$$

$$
Y_{20} = \frac{C_2}{N_{20}\Big(1 + (\frac{N_{21}}{N_{20}})^2 e^{\;\; a_{21}} + (\frac{N_{22}}{N_{20}})^2 e^{\;\; a_{22}}\Big)}, \qquad Y_{02}\frac{C \;\; _2}{N_{02}\Big(1 + (\frac{N_{12}}{N_{02}})^2 e^{\;\; b_{12}} + (\frac{N_{22}}{N_{02}})^2 e^{\;\; b_{22}}\Big)}
$$

$$
\lambda_{ij} = \frac{Y_{i0}N_{0j}e^{b_{ij}}}{Y_{i0}N_{0j}e^{b_{ij}} + Y_{0j}N_{i0}e^{a_{ij}}}, \qquad Y_{ij} = N_{ij}\Big(\frac{Y_{i0}}{N_{i0}}e^{\;\; a_{ij}} + \frac{Y_{0j}}{N_{0j}}e^{\;\; b_{ij}}\Big), \qquad Z_{ij} = \ln\Big(Y_{ij}\sqrt{\frac{\lambda_{ij}(1-\lambda_{ij})}{Y_{i0}Y_{0j}}}\Big)
$$

Then, we build and objective function that consist of the sum of squares of the below equation as a function of $\bar{a}_{i}$, $\bar{a}_{j}$, $\bar{b}_{i}$, $\bar{b}_{j}$

$$
\begin{aligned}
&N_{10} + N_{11}\exp(a_{11}) + N_{12}\exp(a_{12}) = D_{1} \;, \qquad N_{01} + N_{11}\exp(b_{11}) + N_{21}\exp(b_{21}) = D_{1} \\
&N_{20} + N_{21}\exp(a_{21}) + N_{22}\exp(a_{22}) = D_{2} \;, \qquad N_{02} + N_{12}\exp(b_{12}) + N_{22}\exp(b_{22}) = D_{2} \\
&\bar{S}_{i} = \frac{1}{2}(\bar{a}_{i} + \bar{b}_{i}) + \bar{Z}_{i} \qquad \bar{S}_{j} = \frac{1}{2}(\bar{a}_{j} + \bar{b}_{j}) + \bar{Z}_{j}
\end{aligned}
$$

In a special case, when $N_{i} = N_{j} = N$, $D_{i} = D_{i} = D_{i}$, $C_{i} = C_{i} = C_{i}$, $\mu_{i} = \omega_{j}$, the market becomes symmetric and at equilibrium, we have

$$
\begin{aligned}
&N_{10} = N_{01} = N_{20} = N_{02}, \quad N_{12} = N_{21}, \quad N_{11} = N_{22}, \qquad a_{11} = b_{11}, \quad a_{22} = b_{22}, \quad a_{12} = b_{21}, \quad a_{21} = b_{12} \\
&Y_{10} = Y_{01}, \quad Y_{20} = Y_{02}, \quad Y_{12} = Y_{21}, \qquad \lambda_{11} = \lambda_{22} = \frac{1}{2}, \quad \lambda_{12} = 1 - \lambda_{21}
\end{aligned}
$$

Still, the analytic solution of the equilibrium is hard to obtain. If we further assume $D_{1} = D_{2} = D$, we obtain $a_{12} = a_{21}$, $b_{12} = b_{21}$ and $\boldsymbol{\rho}^{a} = \boldsymbol{\rho}^{b}$. Then, by letting $D_{0} = D - N_{10}$, the equilibrium level of variables based on the inputs are

$$
\begin{aligned}
&e^{a_{11}} = \frac{D_{0}}{N_{11}} \frac{1}{1 + e^{\;2(\rho^{Y}_{11} + 2\rho^{a}_{11})}} \qquad e^{a_{12}} = \frac{D_{0}}{N_{12}} \frac{1}{1 + e^{\;2(\rho^{Y}_{12} + 2\rho^{a}_{12})}} \\
&Y_{10} = \frac{N_{0} D_{0}}{N_{0}^{2} D_{0} + N_{11}^{3}(1 + e^{\;2(\rho^{Y}_{11} + 2\rho^{a}_{11})}) + N_{12}^{3}(1 + e^{\;2(\rho^{Y}_{12} + 2\rho^{a}_{12})})} C_{1}, \qquad Y_{20} = Y_{10}\frac{C_{2}}{C_{1}} \\
&Y_{11} = 2\frac{N_{11} Y_{10}}{N_{0}} e^{\;a_{11}}, \quad Y_{22} = \frac{C_{2}}{C_{1}} Y_{11}, \quad Y_{12} = (1 + \frac{C_{2}}{C_{1}}) \frac{N_{11} Y_{10}}{N_{0}} e^{\;a_{11}} e^{\;2\rho^{Y}_{11}} \\
&\lambda_{11} = \lambda_{22} = \frac{1}{2}, \qquad \lambda_{12} = \frac{C_{1}}{C_{1} + C_{2}}, \quad \lambda_{21} = \frac{C_{2}}{C_{1} + C_{2}} \\
&\lambda_{11} Y_{11} = \frac{N_{11} e^{\;a_{11}}}{N_{0}} Y_{10} \qquad \lambda_{12} Y_{12} = \frac{N_{11} e^{\;a_{11}}}{N_{0}} e^{2(\rho^{a}_{11} \;\; \rho_{11})} Y_{10} \\
&\lambda_{22} Y_{22} = \frac{N_{11} e^{\;a_{11}}}{N_{0}} Y_{10} \frac{C_{2}}{C_{1}} \qquad \lambda_{21} Y_{21} = \frac{N_{11} e^{\;a_{11}}}{N_{0}} e^{2(\rho^{a}_{11} \;\; \rho_{11})} Y_{10} \frac{C_{2}}{C_{1}}
\end{aligned}
$$

A.3 Iterative Algorithm for Computing Equilibrium

At an equilibrium, we must determine population and income tables that each of theminclude $(I+1)(J+1)-1$ elements and the sharing rule table with IJ elements. There are $3(I-1)(J-1)$ equations from the AM's components, $I+J$ equations as $N_{i0} + N_{i+} = N_{i}$ and $N_{0j} + N_{+j} = N_{j}$, $I+J$ equations as $N_{i+} = \mu_{i} N_{i}$ and $N_{+j} = \omega_{j} N_{j}$, and $3(I+J)$ equations as (33) to (35). Since $\sum_{i} N_{i+} = \sum_{j} N_{+j}$, $\sum_{i} \bar{a}_{i} = \sum_{j} \bar{a}_{j}$, $\sum_{i} \bar{b}_{i} = \sum_{j} \bar{b}_{j}$, the total number of independent equations is equal to unknown parameters $3IJ + 2(I+J)$.

From

$$
a_{ij} = \rho^{a}_{ij} + \bar{a}_{i} + \bar{a}_{j} - \bar{a} \qquad b_{ij} = \rho^{b}_{ij} + \bar{b}_{i} + \bar{b}_{j} - \bar{b} \tag{42}
$$

we first solve (33) with two $I-1$ and $J-1$ vectors by minimizing the sum of squares of

$$G_i = \frac{1}{D_i}\Big(N_{i0} + \sum_{j=1}^{J} N_{ij}\exp\big(\frac{a_{ij}}{\gamma_{ij}}\big)\Big) - 1 \qquad G_j = \frac{1}{D_{\ j}}\Big(N_{0j} + \sum_{i=1}^{I} N_{ij}\exp\big(\frac{b_{ij}}{\gamma_{ij}}\big)\Big) - 1 \tag{43}$$

using a numerical optimization methods. When the function's arguments are $\bar a_{\ j}$ and $\forall 1 \le i \le I-1 : \bar a_i$ are given, $\bar a_{\ J} = \sum_{i=1}^{I} \bar a_i\ - \sum_{j=1}^{J\ 1} \bar a_{\ j}$ and we obtain

$$a_{ij} = \rho^a_{ij} + \bar a_i\ + \bar a_{\ j} - \frac{1}{I}\sum_{i'=1}^{I} \bar a_{i'}\ , \qquad a_{iJ} = \rho^a_{ij} + \bar a_i\ + \sum_{j=1}^{J\ 1} \bar a_{\ j} + (1 - \frac{1}{I})\sum_{i'=1}^{I} \bar a_{i'}$$

$$G = \sum_{i=1}^{I} G_i^2, \qquad \frac{\partial G}{\partial \bar a_i} = 2\sum_{i=1}^{I} G_i \frac{\partial G_i}{\partial \bar a_i}, \qquad \frac{\partial G_i}{\partial \bar a_{i'}} = \frac{1}{D_i}\sum_{j=1}^{J} \frac{N_{ij}}{\gamma_{ij}}\exp\left(\frac{a_{ij}}{\gamma_{ij}}\right)\frac{\partial a_{ij}}{\partial \bar a_{i'}}$$

$$\frac{\partial a_{ij}}{\partial \bar a_i} = (i = i^0) + (j = J) - \frac{1}{I}$$

The solution gives $\bar a_i$ as a function of $\bar a_{.1}, \ldots, \bar a_{\ J\text{-}1}$. Similarly, solving G_i, when the function's arguments are $\bar a_i$ and $\forall 1 \le j \le J-1, \bar a_{\ j}$ are given yields

$$a_{ij} = \rho^a_{ij} + \bar a_i\ + \bar a_{\ j} - \frac{1}{J}\sum_{j'=1}^{J} \bar a_{\ j'}, \qquad a_{Ij} = \rho^a_{ij} + \bar a_i\ + \sum_{i=1}^{I\ 1} \bar a_i\ + (1 - \frac{1}{J})\sum_{j'=1}^{J} \bar a_{\ j'}$$

$$G = \sum_{i=1}^{I} G_i^2, \qquad \frac{\partial G}{\partial \bar a_{\ j}} = 2\sum_{i=1}^{I} G_i \frac{\partial G_i}{\partial \bar a_{\ j}}, \qquad \frac{\partial G_i}{\partial \bar a_{\ j'}} = \frac{1}{D_i}\sum_{j=1}^{J} \frac{N_{ij}}{\gamma_{ij}}\exp\left(\frac{a_{ij}}{\gamma_{ij}}\right)\frac{\partial a_{ij}}{\partial \bar a_{\ j'}}$$

$$\frac{\partial a_{ij}}{\partial \bar a_{\ j'}} = (j = j^0) + (i = I) - \frac{1}{J}$$

and we can compute $\bar a_{\ j}$ as a function of $\bar a_{1.}, \ldots, \bar a_{I\text{-}1}$. To find similar functions for $\bar b_i$ and $\bar b_{\ j}$, we solve G_j in the same way.

$$b_{ij} = \rho^b_{ij} + \bar b_i\ + \bar b_{\ j} - \frac{1}{I}\sum_{i'=1}^{I} \bar b_{i'}\ , \qquad b_{iJ} = \rho^b_{ij} + \bar b_i\ + \sum_{j=1}^{J\ 1} \bar b_{\ j} + (1 - \frac{1}{I})\sum_{i'=1}^{I} \bar b_{i'}$$

$$G = \sum_{j=1}^{J} G_j^2, \qquad \frac{\partial G}{\partial \bar b_i} = 2\sum_{j=1}^{J} G_j \frac{\partial G_j}{\partial \bar b_i}, \qquad \frac{\partial G_j}{\partial \bar b_{i'}} = \frac{1}{D_{\ j}}\sum_{i=1}^{I} \frac{N_{ij}}{\gamma_{ij}}\exp\left(\frac{b_{ij}}{\gamma_{ij}}\right)\frac{\partial b_{ij}}{\partial \bar b_{i'}}$$

$$\frac{\partial b_{ij}}{\partial \bar b_{i'}} = (i = i^0) + (j = J) - \frac{1}{I}$$

$$b_{ij} = \rho^b_{ij} + \bar b_i\ + \bar b_{\ j} - \frac{1}{J}\sum_{j'=1}^{J} \bar b_{\ j'}, \qquad b_{Ij} = \rho^b_{ij} + \bar b_i\ + \sum_{i=1}^{I\ 1} \bar b_i\ + (1 - \frac{1}{J})\sum_{j'=1}^{J} \bar b_{\ j'}$$

$$G = \sum_{j=1}^{J} G_j^2, \qquad \frac{\partial G}{\partial \bar a_{\ j}} = 2\sum_{j=1}^{J} G_j \frac{\partial G_j}{\partial \bar a_{\ j}}, \qquad \frac{\partial G_j}{\partial \bar a_{\ j'}} = \frac{1}{D_{\ j}}\sum_{i=1}^{I} \frac{N_{ij}}{\gamma_{ij}}\exp\left(\frac{a_{ij}}{\gamma_{ij}}\right)\frac{\partial a_{ij}}{\partial \bar a_{\ j'}}$$

$$\frac{\partial a_{ij}}{\partial \bar a_{\ j'}} = (j = j^0) + (i = I) - \frac{1}{J}$$

Using the these functions, we follow this iterative algorithm by choosing initial values for $\bar a^{(0)}_{.1}, \ldots, \bar a^{(0)}_{\ J\text{-}1}$

and $\bar{b}_{1.},\ldots,\bar{b}^{(0)}_{I\text{-}1}$ such that at iteration $k \geq 1$

1. From (33), find $\bar{a}^{(2k\text{-}2)}_{i}$ as a function of $\bar{a}^{(2k\text{-}2)}_{.1},\ldots,\bar{a}^{(2k\text{-}2)}_{J\text{-}1}$ and $\bar{b}^{(2k\text{-}2)}_{j}$ as a function of $\bar{b}^{(2k\text{-}2)}_{1.},\ldots,\bar{b}^{(2k\text{-}2)}_{I\text{-}1}$

2. compute

$$a^{(2k\text{-}2)}_{ij} = \rho^{a}_{ij} + \bar{a}^{(2k\text{-}2)}_{i} + \bar{a}^{(2k\text{-}2)}_{\ j} - \frac{1}{I}\sum_{i'=1}^{I}\bar{a}^{(2k\text{-}2)}_{i'} \text{ and } b^{(2k\text{-}2)}_{ij} = \rho^{b}_{ij} + \bar{b}^{(2k\text{-}2)}_{i} + \bar{b}^{(2k\text{-}2)}_{\ j} - \frac{1}{J}\sum_{j'=1}^{J}\bar{b}^{(2k\text{-}2)}_{j'}$$

3. compute

$$Y^{(2k\text{-}2)}_{i0} = \frac{C_i}{N_{i0}\Big(1+\sum_{j=1}^{J}(\frac{N_{ij}}{N_{i0}})^{\frac{\gamma_{ij}+1}{\gamma_{ij}}}\exp\big(-\frac{a^{(2k\text{-}2)}_{ij}}{\gamma_{ij}}\big)\Big)}$$
$$Y^{(2k\text{-}2)}_{0j} = \frac{C_{\ j}}{N_{0j}\Big(1+\sum_{i=1}^{I}(\frac{N_{ij}}{N_{0j}})^{\frac{\gamma_{ij}+1}{\gamma_{ij}}}\exp\big(-\frac{b^{(2k\text{-}2)}_{ij}}{\gamma_{ij}}\big)\Big)}$$

4. compute $\boldsymbol{\lambda}^{(2k\text{-}2)}$ from (11), and couples income from

$$Y^{(2k\text{-}2)}_{ij} = N^{1/\gamma_{ij}}_{ij}\Big(\frac{Y^{(2k\text{-}2)}_{i0}}{N^{1/\gamma_{ij}}_{i0}}\exp(-\frac{a^{(2k\text{-}2)}_{ij}}{\gamma_{ij}}) + \frac{Y^{(2k\text{-}2)}_{0j}}{N^{1/\gamma_{ij}}_{0j}}\exp(-\frac{b^{(2k\text{-}2)}_{ij}}{\gamma_{ij}})\Big) \tag{44}$$

5. compute $Z^{(2k\text{-}2)}_{ij} = \gamma_{ij}\ln\Big(Y^{(2k\text{-}2)}_{ij}\sqrt{\frac{\lambda^{(2k\text{-}2)}_{ij}(1-\lambda^{(2k\text{-}2)}_{ij})}{Y^{(2k\text{-}2)}_{i0}Y^{(2k\text{-}2)}_{0j}}}\Big)$

6. find

$$\bar{a}^{(2k\text{-}1)}_{i.}\big(\bar{a}^{(2k\text{-}2)}_{.1},\ldots,\bar{a}^{(2k\text{-}2)}_{J\text{-}1},\bar{b}^{(2k\text{-}2)}_{1.},\ldots,\bar{b}^{(2k\text{-}2)}_{I\text{-}1}\big) = 2\bar{S}_{i} - 2\bar{Z}^{(2k\text{-}2)}_{i} - \bar{b}^{(2k\text{-}2)}_{i}$$
$$\bar{b}^{(2k\text{-}1)}_{j}\Big(\bar{a}^{(2k\text{-}2)}_{.1},\ldots,\bar{a}^{(2k\text{-}2)}_{J\text{-}1},\bar{b}^{(2k\text{-}2)}_{1.},\ldots,\bar{b}^{(2k\text{-}2)}_{I\text{-}1}\Big) = 2\bar{S}_{\ j} - 2\bar{Z}^{(2k\text{-}2)}_{\ j} - \bar{a}^{(2k\text{-}2)}_{\ j}$$

7. From (33), find $\bar{a}^{(2k\text{-}1)}_{\ j}$ as a function of $\bar{a}^{(2k\text{-}1)}_{1},\ldots,\bar{a}^{(2k\text{-}1)}_{I\text{-}1}$ and $\bar{b}^{(2k\text{-}1)}_{i}$ as a function of $\bar{b}^{(2k\text{-}1)}_{.1},\ldots,\bar{b}^{(2k\text{-}1)}_{J\text{-}1}$,

8. compute $\boldsymbol{a}^{(2k\text{-}1)},\boldsymbol{b}^{(2k\text{-}1)},\boldsymbol{Y}^{(2k\text{-}1)}_{0},\boldsymbol{Y}^{(2k\text{-}1)}_{0},\boldsymbol{\lambda}^{(2k\text{-}1)},\boldsymbol{Y}^{(2k\text{-}1)}$ and $\boldsymbol{Z}^{(2k\text{-}1)}$ similar to steps 2 to 5.

9. find

$$\bar{a}^{(2k)}_{.j}\big(\bar{a}^{(2k\text{-}1)}_{1},\ldots,\bar{a}^{(2k\text{-}1)}_{I\text{-}1},\bar{b}^{(2k\text{-}1)}_{1},\ldots,\bar{b}^{(2k\text{-}1)}_{J\text{-}1}\big) = 2\bar{S}_{\ j} - 2\bar{Z}^{(2k\text{-}1)}_{\ j} - \bar{b}^{(2k\text{-}1)}_{\ j}$$
$$\bar{b}^{(2k)}_{i}\Big(\bar{a}^{(2k\text{-}1)}_{1},\ldots,\bar{a}^{(2k\text{-}1)}_{I\text{-}1},\bar{b}^{(2k\text{-}1)}_{1},\ldots,\bar{b}^{(2k\text{-}1)}_{J\text{-}1}\Big) = 2\bar{S}_{i} - 2\bar{Z}^{(2k\text{-}2)}_{i} - \bar{a}^{(2k\text{-}2)}_{i}$$

A.4 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, MARST, 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.

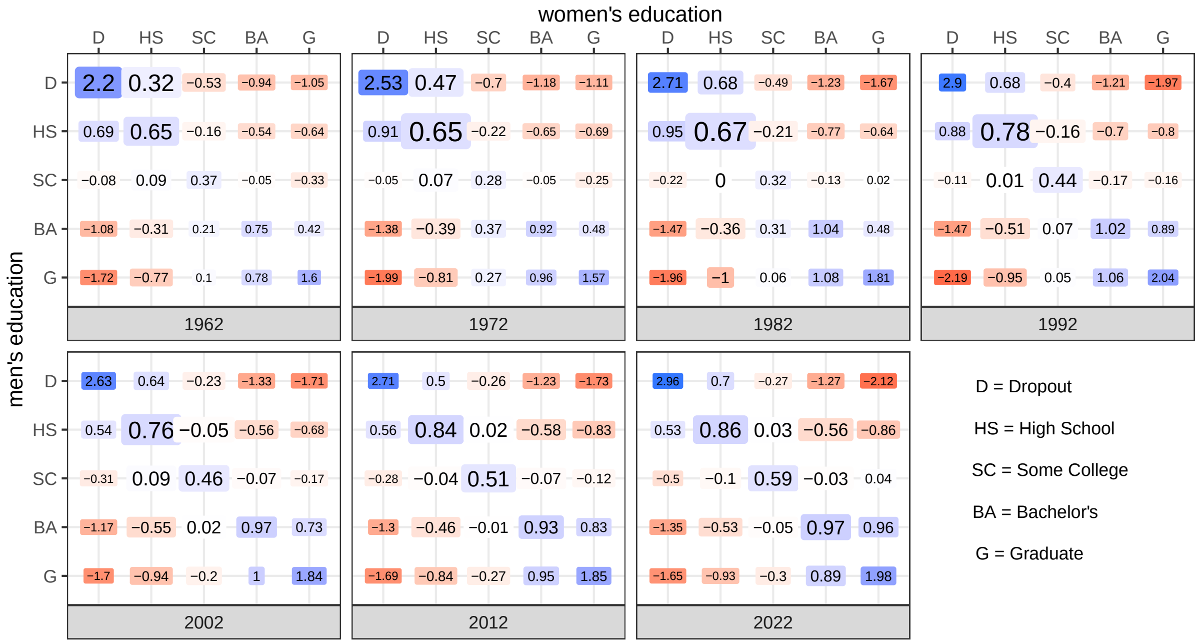

Figure 18: AM in population measured by geometric mean of log odds ratios. The size of the boxes in the figure is proportional to the product of all four elements of the corresponding odds ratio in (1). Notably, in all years, the values of diagonal elements are positive, while the values of anti-diagonal 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.

- 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).

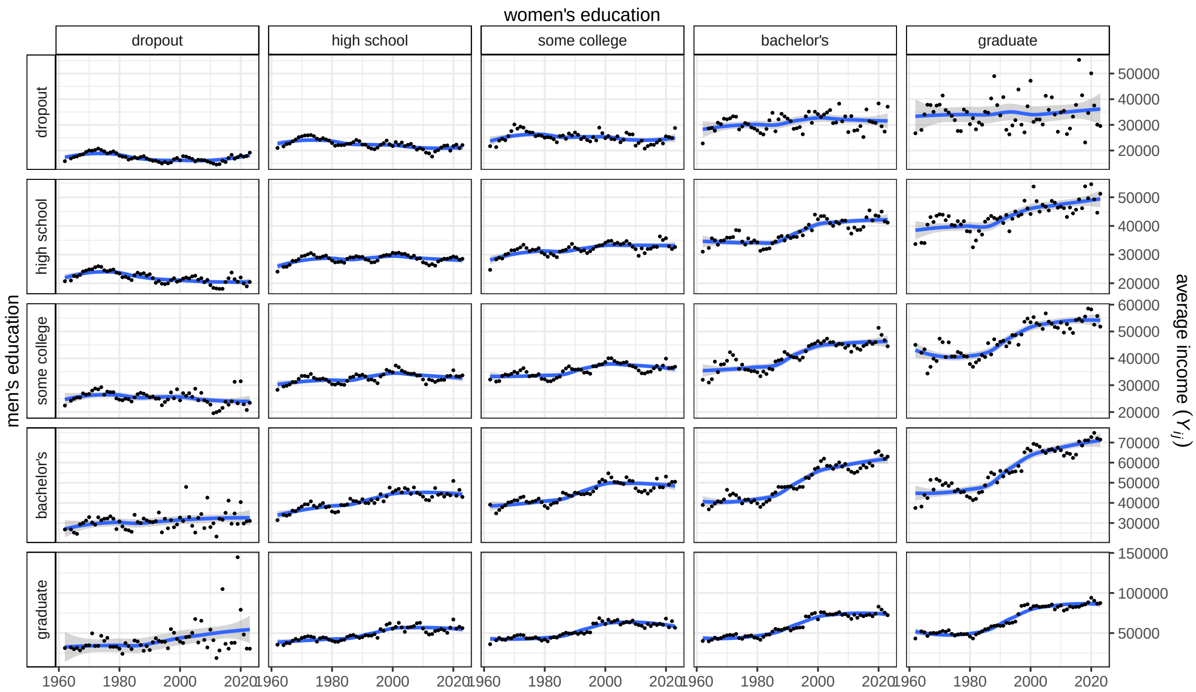

Figure 19: Average income and its LOESS estimation

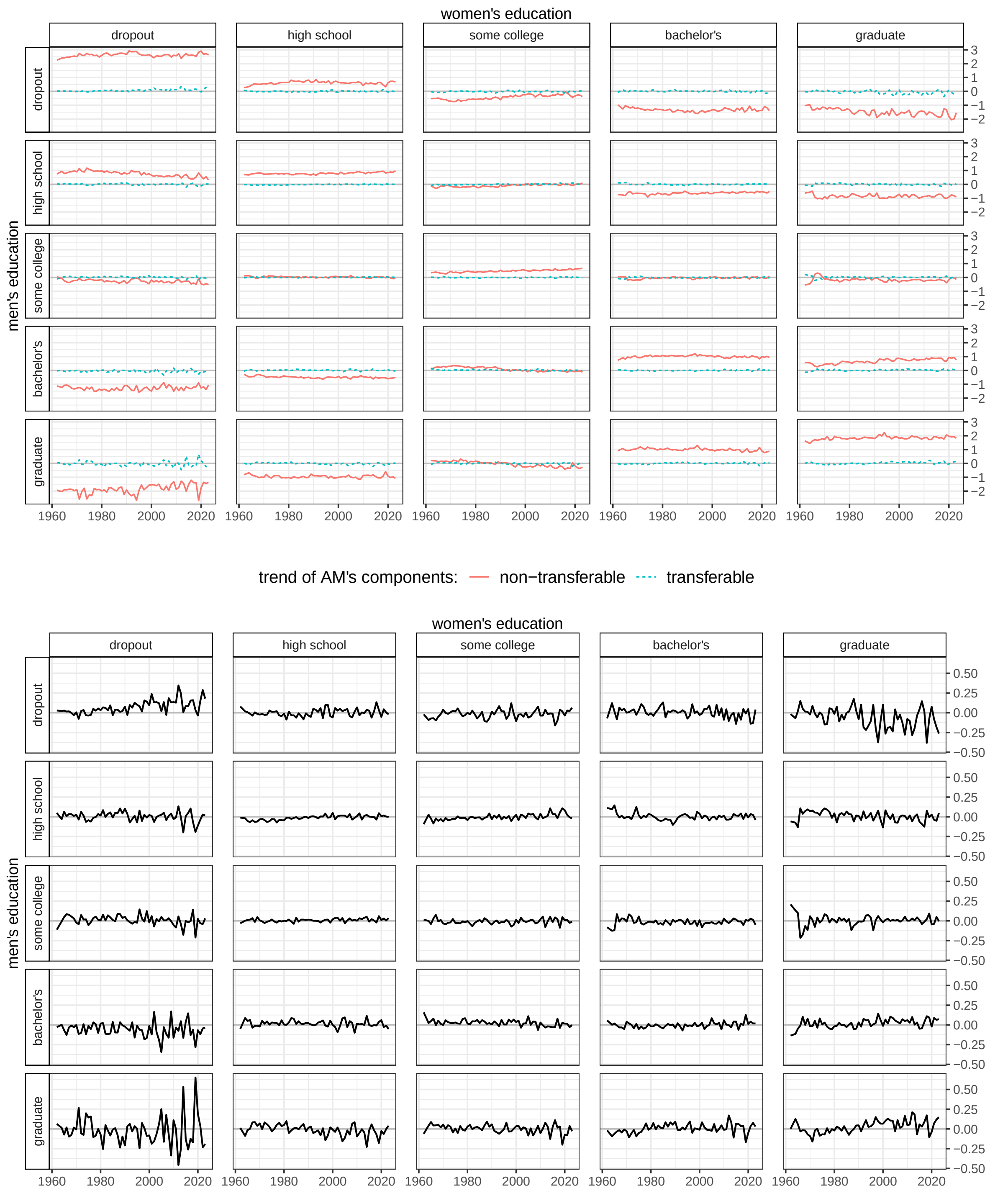

Figure 20: The trends of non-pecuniary ($\frac{\boldsymbol{\rho}^a+\boldsymbol{\rho}^b}{2}$) and pecuniary ($\boldsymbol{\rho}^Y$) components of AM as defined in (24). For better illustration, only the trend of pecuniary component is depicted in the bottom graph.

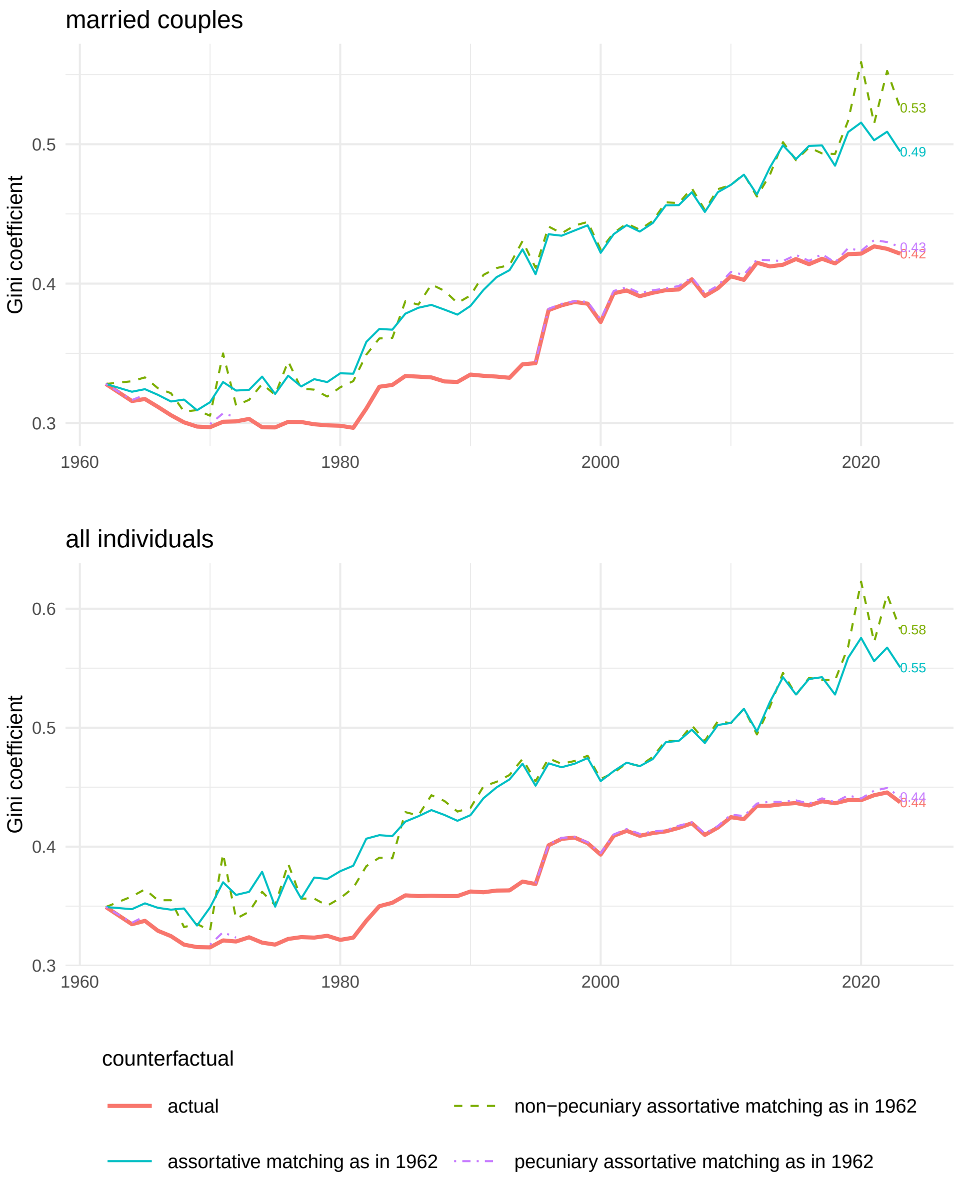

Figure 21: Counter-factual inequality with $\lambda_{ij} = 0.25 + 0.5\lambda_{ij}$. The top graph is at household level and in the bottom graph married couple of each type are counted as two individual with income according to the sharing rule λ_{ij}.