FUNDAÇÃO GETULIO VARGAS
ESCOLA de PÓS-GRADUAÇÃO em ECONOMIA

Diogo Wolff Surdi

Marriages, Educational Attainment and Intergenerational Mobility

Rio de Janeiro

2024

Diogo Wolff Surdi

Marriages, Educational Attainment and Intergenerational Mobility

Dissertação para obtenção do grau de mestre apresentada à Escola de Pós-Graduação em Economia

Área de concentração: Microeconomia Aplicada

Orientador: Carlos Eugênio Ellery da Costa

Rio de Janeiro

2024

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pelo Sistema de Bibliotecas/FGV

Surdi, Diogo Wolff

Marriages, Educational Attainment and Intergenerational Mobility / Diogo Wolff Surdi. – 2024.

38 f.

Dissertação (mestrado) – Escola Brasileira de Economia e Finanças.
Orientador: Carlos Eugênio Ellery da Costa.
Inclui bibliografia.

1. Orçamento familiar. 2. Renda - Distribuição. 3. Casamento. 4. Mobilidade social. I. Costa, Carlos Eugênio da. II. Fundação Getulio Vargas. Escola Brasileira de Economia e Finanças. III. Título.

CDD – 332.024

Elaborada por Marcelle Costal de Castro dos Santos – CRB-7/016/20

FUNDAÇÃO GETULIO VARGAS
MESTRADO EM ECONOMIA
EPGE ESCOLA BRASILEIRA DE ECONOMIA E FINANÇAS - FGV EPGE

DIOGO WOLFF SURDI

"MARRIAGES, EDUCATIONAL ATTAINMENT AND INTERGENERATIONAL MOBILITY''.

DISSERTAÇÃO APRESENTADO/A AO CURSO DE MESTRADO EM ECONOMIA PARA OBTENÇÃO DO GRAU DE MESTRE (A) EM ECOOMIA.

ESTE É UM TRABALHO ORIGINAL ONDE FOI VERIFICADA A NÃO EXISTÊNCIA DE PLÁGIO E DE UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL, NÃO EXPLICITADA, NO CORPO DO TRABALHO.

DATA DA DEFESA: 27/03/2024

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

PRESIDENTE DA COMISSÃO EXAMINADORA: PROFº/ª CARLOS EUGÊNIO ELLERY LUSTOSA DA COSTA

<ASSINADO ELETRONICAMENTE>
PROFº/ª CARLOS EUGÊNIO ELLERY LUSTOSA DA COSTA
ORIENTADOR

<ASSINADO ELETRONICAMENTE>
PROFº/ª MAURO RODRIGUES JR
MEMBRO

<ASSINADO ELETRONICAMENTE>
PROFº/ª ANDREA MARÍA FLORES
MEMBRO

RIO DE JANEIRO, 27 DE MARÇO DE 2024.

<ASSINADO ELETRONICAMENTE>
PROFº LUCAS JÓVER MAESTRI
COORDENADOR

<ASSINADO ELETRONICAMENTE>
PROFº ANTONIO DE ARAUJO FREITAS JUNIOR
PRÓ-REITOR DE ENSINO, PESQUISA E PÓS-GRADUAÇÃO

Para Laura

Agradecimentos

Agradeço ao meu orientador Carlos Eugênio, por elucidar direções para meu trabalho; A diversos professores da EPGE por comentários e sugestões ao longo do mestrado, em particular Andrea Flores; Aos grupos de economia da família e de macroeconomia da EPGE; ao CDMC, que tornou esta jornada possível; à minha família, por me apoiar durante toda minha vida; e à Júlia, minha companheira e melhor amiga.

Resumo

Um modelo de gerações sobrepostas é utilizado para quantificar o efeito que o nível de ordenação de casais de acordo com a educação tem sobre a desigualdade e a persistência de renda. Utilizando microdados dos EUA e dados multigeracionais, o modelo é calibrado para a economia americana do começo dos anos 2000. O grau de ordenação de casais é um fator significativo para a persistência de renda prevista pelo modelo. Ademais, efeitos de equilíbrio do mercado de casamentos se demonstram importantes para análises contrafactuais.

Palavras-chave: Casamento, Desigualdade, Economia de domicílio

Abstract

An overlapping generations model is deployed to quantify the effect that the level of sorting of couples with respect to education has on income inequality and persistence. Using US microdata and intergenerational surveys, the model is calibrated to the US economy in the early 2000s. Assortative matching is shown to be a significant factor in the persistence of income predicted by the model. Furthermore, marriage market equilibrium effects cannot be dismissed in counterfactual analyses.

Keywords: Marriage, Inequality, Household Economics

List of Figures

1 Identification of labor supply in equilibrium xvi
2 NLSY data xix
3 Degree of assortativeness of each 5-year US cohort xxi
4 Fertility rates of adults born in the 1960s xxii
5 Persistence measures as a function of shocked sorting parameters xxvi
6 Effects of a shift in available hours for women on income inequality xxvii
7 Effect of changes in returns to marriage on income persistence xxviii
8 Measure of couple types as a function of the shock to available hours to women . xxxiii
9 Covariance of income of fathers and boys xxxiv
10 Covariance of income of mothers and daughters xxxv
11 Measure of couple types as a function of the shock to college costs of men xxxvi

List of Tables

1 Marriage patterns, early 60s cohort xx

2 Wages, CPS 2003 cohort xxii

3 Calibration of the second stage xxiv

4 Marriage rates and early education patterns xxiv

5 Cost function estimates xxv

6 Model estimates of selected income statistics, imposing steady-state xxv

Contents

1 Introduction xi

2 Model xiii

3 Equilibrium and dynamics xvii

4 Data xix

5 Estimation xxii

6 Results xxiii

6.1 Comparative statics xxv

7 Counterfactuals xxvi

8 Conclusion xxix

1 Introduction

Families play a fundamental role in the development of every person. Throughout most of our youth, it is our parents who decide our activities, the goods we consume, the quality of the school in which we are enrolled, and many other factors that shape our tastes and abilities. As these choices can be correlated with their own background, it is reasonable to expect that the success of their children is also correlated with theirs.

As a significant share of households are composed of couples with children, the aforementioned parental background is the joint history of both parents, and their individual decisions could have complementary effects toward the development of their children. Given the general hypothesis that parents care for their children, these effects generate an incentive for parents to live together as opposed to raising children as singles.

In order for a child to be born, their parents must have met beforehand. Such meetings are not random, and the societal patterns which induce them define which types of family (couples) are more likely to appear. The distribution of households that are formed determines the distribution of children in each parental background, which affects in turn the inequality of the next generation and the persistence of income across generations.

In this paper, our aim is to uncover the strength that these patterns have on such income measures by defining a marriage market through which individuals match. In each generation, young adults participate in a two-stage game, where they first choose their education level, and then match in a frictionless environment. After the marriage market clears, each household draws the number of children they will have and then chooses their time expenditure, which is split between leisure, labor, and time spent with children. The last variable, interpreted as childcare, is the factor that determines the costs of reaching a higher level of education that their children will face.

Building on this framework, we construct a multi-generational economy, allowing us to study long-run effects. The model is calibrated to US data, targeting the population of adults born around 1960, replicating their marriage patterns and time expenditures. As the determinants that define education costs have shifted substantially through recent decades, we calibrate their values to endogenize the measure of individuals in this cohort in each education bracket as the steady-state of the economy. The main result is that the persistence of income between boys and their fathers (girls and their mothers) is estimated at 6.6% (20.5%).

Given our results, we employ several exercises that shed light on multiple ways in which endogenous matching impacts the results, a channel that is largely untouched in the macroeconomics literature. By shocking the incentives to sort positively, we show that an increase of 1% in the measure of couples with the same education level is associated with a 3.5% (2%) increase in the income persistence measure of men (women). To put this measure in perspective, the measure of couples in this group has increased 5% from generations born in 1958 to 1978.

Finally, we shut down different stages of the decision of individuals to highlight how each element plays a role in the model. Fixing the education choices, we show the effects of a counterfactual decrease in the exogenous homework hours that women must employ; The shift in the Gini index calculated under this constraint is significantly greater than that under endogenous education. On the other hand, fixing the perceived gains to marriage of each individual, we increase the fixed costs of obtaining a college degree for men; This results in mixed effects, with the shift in income persistence of men being hampered with respect to the complete equilibrium, while it is heightened for women.

The contributions of this project are at the intersection of family economics and macroeconomics with heterogeneous agents. The study of marriage sorting and income inequality dates back a few decades, with prominent works such as Aiyagari et al. (2000), Fernandez & Rogerson (2001), Greenwood et al. (2003), and Fernandez et al. (2005). In this paper, we focus on income persistence, a variable that is underexplored in these articles. This is done in a setting that embeds new techniques from family economics, which allow us to overcome the difficulties of the aforementioned papers.

In the more recent literature, works such as Greenwood et al. (2016) and Pilossoph & Lin Wee (2023) link labor market decisions and the marriage market, accounting for a significant share of household-level inequality. By conducting our analysis with multiple generations, we extend this line of research to the measurement of inequality of children's outcomes with respect to marital patterns of their parents' generation, expanding the magnitude of the effects previously measured.

Furthermore, the role of assortativeness in educational attainment has received increased attention in recent years, as shown by Chiappori et al. (2017, 2018, 2020). The inclusion of childcare patterns is an extension of such models that affects the returns to education, as better educated spouses make for more efficient parents. The effects of this addition can further clarify the education decisions and the role of marital gains in these decisions.

The structure of the paper is as follows: in Section 2, the model is introduced; in Section 3, the equilibrium of the model is defined, and so is the structure of stationary distributions; Section 4 details the datasets used, and Section 5 determines the details of estimation; Results of the estimation and counterfactuals are shown in Sections 6 and 7; Section 8 concludes the paper.

2 Model

To understand the mechanics that we aim to replicate, let us look at the lifetime of a woman in the model economy. As a child, she lives inside a household with either one or two adult parents and a number of siblings that may range from zero to three. Her parents decide the total amount of childcare Q that they provide to all children, so she does not make any decisions within this household. There is no heterogeneity in this stage of the model.

In the next period, the woman participates in a two-stage game, whose final outcome is the type of household she will compose as an adult - note that we distinguish adult from parent, as some households may have no children. Given her education level, a value $H_w \in H = \{HS : \text{ high school or less}, \; SC : \text{ some college}, \; CP : \text{ college and above}\}$, there are four possible households for her to choose: a marriage with a man of any of the three education levels above, whose education we will call H_m, or the option to remain single.

Each of these options generates a certain utility value for the woman, intrinsically tied to the economic model we will present later. This value is a result of the equilibrium of a frictionless marriage market in which all individuals who became adults in this period are either matched into new married households or end up as singles. It is independent of the specific man she marries, being conditional only on his education level, and it is an expected value, as the number of children she will have is an exogenous random variable that is drawn after households are formed. Furthermore, there is a random component to the utility of each choice, whose value is revealed before making a choice.

Given the set of utility values, we can recover the expected utilities of the woman for all of her education levels, before the random components of her utility are revealed. These expectations are then used to determine the best level of education to attain, which is the first stage of the game. The childcare state Q enters the game through a linear cost function, which penalizes the value of each education level and whose coefficients are specific to each level and gender. As is the case of the second stage, there is a random component which shifts the cost function as well, allowing the model to comport heterogeneity in behavior.

After the game is completed, the household where the woman resides births a random number of children. The distribution of this random variable is conditional on the education level of both spouses, in the case of married couples, and on gender, in the case of singles. The family decides then how to allocate the available hours of each adult in the household, where the budget is conditional on exogenous housework, which is specific to each education level and gender. The available hours are divided into leisure, market work, and time spent with children. In the case of couples, the time spent with children of each spouse is combined to form total childcare, a public good of the household.

Let us then describe the household model. Suppose each individual has the following utility function:

$$U(Q, c, l) = Q^{\alpha}(\lambda c^{\sigma} + (1 - \lambda)(Hl)^{\sigma})^{\frac{1}{\sigma}}$$

Where Q is a public good of the household, interpreted as early education of children, c is private consumption, H is human capital, and l is leisure time. As each unit of leisure costs H (which also equals wages), the utility function with Hl as an argument makes all individuals treat leisure as if it had the same cost of the consumption good. This implies that individuals with different levels of human capital, but equal Q, have the same marginal gain to income.

To see this, suppose that the decisions of couples are Pareto efficient. Then, we can view the household-level problem as a two-stage process, where first Q is chosen, and then each partner is assigned a share ρ of the remaining income of the household. Given that a man with education H_m receives income share ρ_m, his indirect utility is:

$$v_m(Q, \rho_m, H_m) = Q^{\alpha} \left(1 - \lambda + \lambda \left(\frac{\lambda}{1 - \lambda}\right)^{\frac{\sigma}{1-\sigma}}\right)^{\frac{1}{\sigma}} \frac{\rho_m}{1 + \left(\frac{\lambda}{1-\lambda}\right)^{\frac{1}{1-\sigma}}}$$

This is entirely independent of H_m. Now, suppose that the good Q is produced with time use of parents, namely, $Q = (H_m t_m)^{\beta} (H_w t_w)^{1-\beta}$. For the household allocation to be Pareto efficient, it must maximize a program of the following form

$$\max_{t_m, t_w, \rho_m} \left((H_m t_m)^{\beta}(H_f t_w)^{1-\beta}\right)^{\alpha} K(\rho_m + \psi(H_m(1 - t_m) + H_f(1 - t_w) - \rho_m))$$

where K is a constant, and the weights of each individual are 1 and ψ. The solution can only be interior if $\psi = 1$, therefore, these preferences have the property of transferable utility.

Under the assumption of allowing only interior solutions, the complete problem of a household formed by individuals of types (H_m, H_w) is:

$$\max_{c_m,c_w,l_m,l_w,t_m,t_w} \left((H_m t_m)^{\beta}(H_f t_w)^{1-\beta}\right)^{\alpha} \left((\lambda c_m^{\sigma} + (1-\lambda)(H_m l_m)^{\sigma})^{\frac{1}{\sigma}} + (\lambda c_w^{\sigma} + (1-\lambda)(H_f l_w)^{\sigma})^{\frac{1}{\sigma}}\right)$$
$$\text{s.t.} \quad c_m + c_w + H_m l_m + H_f l_w = H_m(\tau(H_m) - t_m) + H_f(\tau(H_f) - t_w)$$

The couple chooses their time allocation jointly, and each spouse has to use a share $1 - \tau(H)$ of their time for home production, which is exogenous. In the case of single households, the production function is replaced by $Q = (Ht)^{\xi}$. The solution to the couple's problem is summarized as follows:

$$H_w t_w = \frac{\alpha(1-\beta)}{1+\alpha}(H_m\tau(H_m) + H_w\tau(H_w)) \tag{1}$$
$$H_m t_m = \frac{\alpha\beta}{1+\alpha}(H_m\tau(H_m) + H_w\tau(H_w)) \tag{2}$$
$$H_m l_m + H_w l_w = \frac{H_m\tau(H_m) + H_w\tau(H_w)}{1+\alpha} \frac{1-\lambda}{\left(1-\lambda+\lambda\left(\frac{\lambda}{1-\lambda}\right)^{\frac{\sigma}{1-\sigma}}\right)} \tag{3}$$
$$c = \left(\frac{\lambda}{1-\lambda}\right)^{\frac{1}{1-\sigma}} Hl \tag{4}$$

The childcare choices and total income of households are identified, but the labor supply of each spouse is not. This is a result of the equal weights that the household places on the welfare of either spouse. As both private goods (c and l) are affine functions of the total utility of the household, we can recover the labor supply using the sharing rule $\mu(H_m, H_w)$ extracted from the equilibrium of the marriage market. This is exemplified in Figure 1.

The blue line determines all possible consumption bundles that split the function of total value of marriage $G(\Gamma)$. The dashed red lines are the augmented leisure levels that equate married utility to the utility of a single man ($U(H_m)$) and woman ($V(H_w)$), written as inverse functions. The intercepts a and b define the minimum amount of leisure that each spouse must have to prefer marriage (ex ante to the random utility shock), a condition that must be imposed to have stable marriages. The sharing rule defines the weights of the linear combination of the two points in equilibrium, identifying both l_m and l_w.

If the total measure of a marriage type is smaller than the measure of singles of that type for a gender, the results from Choo and Siow (2006) imply that the economic surplus allocated

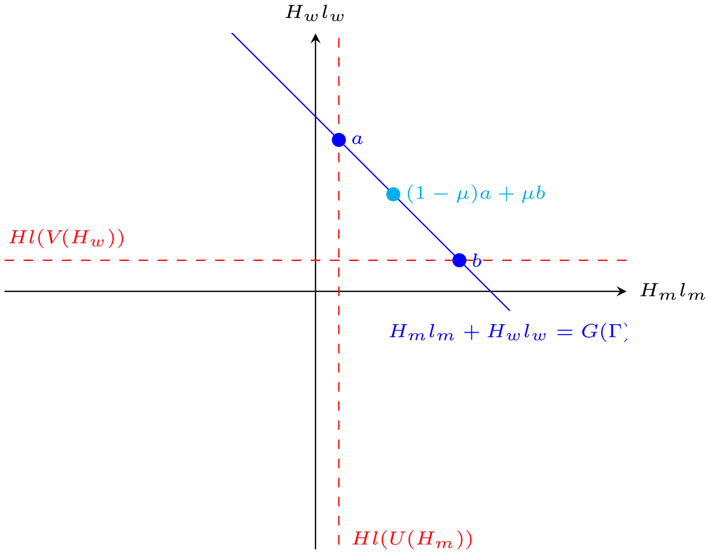

Figure 1: Identification of labor supply in equilibrium. Axes represent augmented leisure of each spouse. The dashed lines are the allocations equivalent to the utility of being single, while the solid line represents all possible allocations in marriage. The pale blue dot is the choice in the marriage market equilibrium.

in equilibrium to that spouse must be negative, with such marriages only occurring when they receive a large enough draw of the random utility shock. In our context, this is equivalent to a nonconvex sharing rule, which does not violate the model, but can generate negative amounts of leisure. In this case, the set of parameters is considered unfeasible, with important consequences for the existence of equilibrium.

Having described the household, we may proceed to the game. The decision on the second stage has the following form:

$$\max_{H_m \in \{H, \emptyset\}} (1-\mu(H_m, H_w))\Phi(H_m, H_w) + \varepsilon_{H_m}$$

Where $\Phi(H_m, H_w)$ is the expected surplus generated by marriage between individuals of type H_m and H_w, $\mu(H_m, H_f)$ is the share of that surplus that is allocated to the husband in equilibrium, and ε_{H_m} is the preference shock, which is unique to each woman, but independent of the specific man she marries, indexed only by his level of education. Given a type-1 extreme-value distribution for each ε_H, the ex-ante probability of choosing to marry a partner type is:

$$P(H_m|H_w, \mu, \Phi) = \frac{exp((1-\mu(H_m, H_w))\Phi(H_m, H_w))}{\sum_{H'_m} exp((1-\mu(H'_m, H_w))\Phi(H'_m, H_w))} \tag{5}$$

Maximization for men is similar, receiving instead a share $\mu(H_m, H_w)$ of the marital surplus.

The marriage market reaches its equilibrium through the variation of the matrix of shares μ, which act as the price of marrying a partner of any given educational level, and the equilibrium concept will be described in more detail in the next section.

This is a version of the standard model of Choo and Siow (2006), and we can derive in equilibrium the average excess utility of each type of individual, with respect to being single. Define these values as $\bar{U}(H_m)$ for men and $\bar{V}(H_w)$ for women, and let μ_{xy} be the measure of marriages between types x and y, and m_x (w_y) be the measure of men (women) of type x (y). We have then the following formulas:

$$\bar{U}(H) = -\log\left(\frac{\mu_H}{m_H}\right)$$
$$\bar{V}(H) = -\log\left(\frac{\mu_H}{w_H}\right)$$

These values determine the expected economic gain of an individual as a function of their education level, and enter the individual problem through the first stage of the game.

Looking now at the first stage, each individual enters the game with an amount of early education Q, which is mapped to the costs $c_H(Q)$ of each level of education. Define $U(H_m)$ ($V(H_w)$) as the utility of being a single man (woman) with education H_m (H_w). The problem of utility maximization in this stage for women is to choose an education level as follows:

$$\max_{H_w \in H} V(H_w) + \bar{V}(H_w) - c_{H_w}(Q) + \nu_{H_w} \tag{6}$$

Where ν_{H_w} is a random shock associated with each education level, which is type-1 extreme-value distributed. The ex ante probability of each education choice follows the same form as equation 5, generating a function of Q.

In order to calculate the equilibrium of the model, we need then only three objects: the surplus matrix $\Phi(H_m, H_w)$, the distribution of children in each level of early education Q, and the cost functions $c_H(Q)$.

3 Equilibrium and dynamics

The solution concept for an equilibrium in the game is similar to that found in Chiappori et al. (2018). Suppose ε_{H_f} (ε_{H_m}) is the unobserved utility a man (woman) receives from marrying a woman (man) of type H_f (H_m). We impose the standard assumption of separability from

Choo and Siow (2006):

Separability: *Given the surplus matrix* $\Phi(H_m, H_f)$ *and individuals with types* (H_m, H_f) *and shocks* $(\varepsilon_{H_f}, \varepsilon_{H_m})$*, the joint utility of marriage must be:*

$$\tilde{\Phi}(H_m, H_f) = \Phi(H_m, H_f) + \varepsilon_{H_f} + \varepsilon_{H_m}$$

Furthermore, the utility of singles is composed only of the unobservable component $\varepsilon_{H_\emptyset}$.

Given this assumption, it is known (see Chiappori & Salanié (2020) for a proof) that there exists a stable matching for the problem. This result means that we can find a pair of matrices U and V that split the surplus for all types of couples, and the entries of the matrices are the amount of the observed surplus that each spouse receives.

With respect to the existence of an equilibrium, the game defined in the model is equivalent to that of the model of Chiappori et al. (2018), in the sense that the last stage generates a specific number (namely, the utility conditional on the results of education and marriage), which is what the players respond to. The existence is guaranteed as their proof still holds; however, the equilibrium might be unfeasible, as stated in the previous section.

In this context, an invariant distribution can be defined as a measure of households such that the total number of men and women in each education bracket matches that of the previous generation. Under homogeneous fertility, with couples having two children and singles having one, it can be shown that an invariant distribution always exists (see Appendix A for a proof). The proof does not hold with heterogeneous fertility, and so far it is unclear under what conditions an invariant distribution should exist.

To overcome this issue, we assume that the distribution of individuals of the data is itself an invariant distribution, determining the education costs as to make this true. This assumption allows us to fully characterize the model as a snapshot of the economy in the early 2000s, imposing the education costs faced by all generations as constant over time. The drawback is that we necessarily will not match the changes in education choices documented in the past few decades (the actual children's generation), but we can still discuss the transition probabilities that the model generates, and compare them to the data.

4 Data

As children and parents participate in the marriage markets in different periods of time, the data sets used for estimation must be carefully managed. The starting point is the NLSY79, which contains information on parental background of children born between 1957 and 1964, and tracks their outcomes throughout life. Of these observations, we will keep all men born in 1959-1963 and women born in 1960-1964, and we suppose that all education has been finished when an individual is 30 years old. We then collect the choices that will be used in the estimation of the education costs by calculating the proportion of each level of education as a function of the education of the parents of these children. The proportions of college attainment are shown in Figure 2.

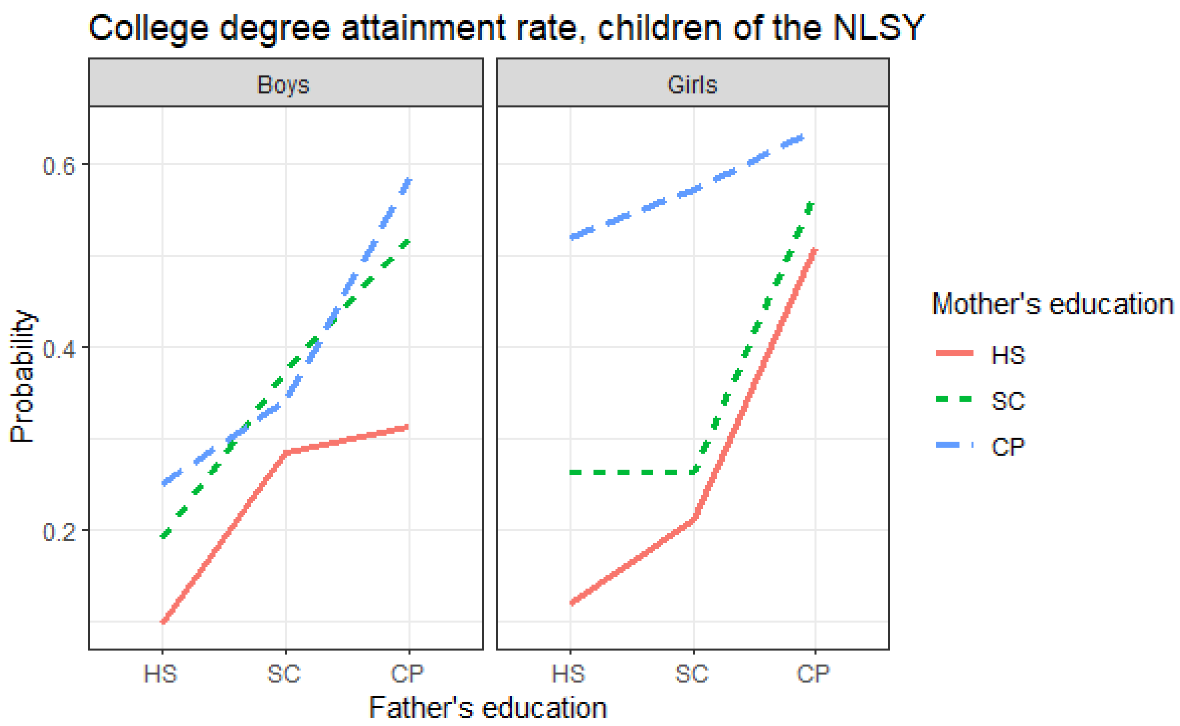

Figure 2: NLSY data

The NLSY data shows that children whose parents have both a higher level of education are more likely to attain a college degree, though we cannot assert that there are complementarities, as the slopes of the lines in the graph do not generally increase from $HS - SC$ to $SC - CP$.

To calculate the proportions of marriages over time, we define a cohort as all men born in a 5-year period and all women born in the same period, shifted one year forward, as a large share of marriages is composed of wives a year younger than their husbands. The data come from ACS waves 2006 through 2019, including only households that are either legally married

Table 1: Marriage patterns, early 60s cohort. Each cell contains the measure of households of each type, using the data from recent ACS waves. Columns represent the education level of women, and rows of men.

	High-School	Some College	College+	Single men
High-School	0.1648	0.0668	0.0268	0.1184
Some College	0.0514	0.0901	0.0484	0.0756
College+	0.0236	0.0601	0.2011	0.0731
Single women	0.0994	0.0816	0.0860	

or both contain a head and a nonmarried partner of the head. From this subset, couples with more than one marriage for either spouse and couples with a spouse still in school are removed. We consider a man (woman) to be single if he reaches 40 (39) years of age without being in a relationship, either as a head or as a partner to a head.

This data set highlights how the sorting channel may have become more prevalent over time. To assess trends in assortativeness in the US, we employ the SEV index by Chiappori et al. (2020), which is 0 under random matching and ∞ under perfect positive sorting; therefore, positive values for this index imply positive assortativeness between the classes considered. As the SEV index is defined for 2×2 matrices, we calculate its value for the three pairs $\{(HS, CP), (HS, SC), (SC, CP)\}$.

In Figure 3, we plot the estimates of this index over time, using 5-year cohorts of individuals born from 1928 to 1978. The first important fact to note is that all index values are positive, indicating that, for any given pair of education levels and periods, individuals match those similar to them. Furthermore, we can see that, especially in the more recent waves, all indices seem to be increasing, evidencing a strengthening of sorting in the economy.

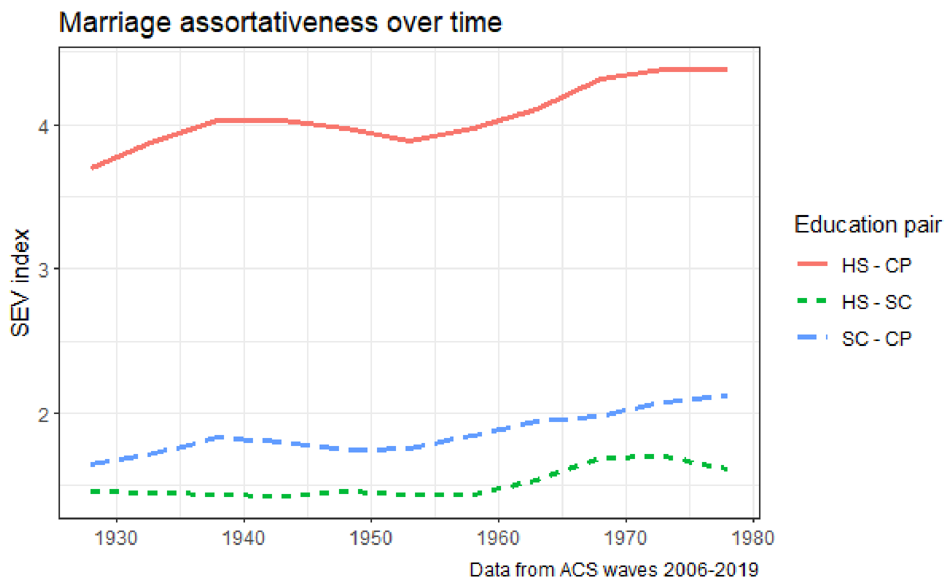

Figure 3: Degree of assortativeness of each 5-year US cohort. The measure is specific to each pair of classes, and more positive values indicate a higher degree of assortativeness.

The remaining datasets generate the moments that are required to estimate the choices of each type of couple, namely, the time use decisions, and the market wage that each agent faces. Time use is calculated based on the American Time Use Survey (ATUS), which has rich information on the time expenditure of individuals in the US. The 2003 wave of the ATUS is used, as in this year the marriage market defined above is finished.

We suppose that each individual has 112 weekly hours available, accounting for 8 hours of sleep per day, and that they must split between market work, child care, and leisure. These hours are discounted by home production, considered exogenous, which is also estimated using ATUS. As ATUS only interviews one individual in each household, we cannot jointly estimate the time spent with children for spouses of each human capital pair, therefore, all moments used are conditional on the education of both spouses but do not reflect complete households. All individuals kept in the sample are between 39 and 45 years old.

To estimate market wages, we use the CPS 2003 wave. The following regression is run:

$$\log(wage_i) = \beta_0 + \beta_1 Educ_i \times gender_i + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 race_i + \zeta_i + \varepsilon_i$$

Where $wage$ is the income from the wages of an individual, $Educ$ is their education level as classified by our cases, and ζ is a fixed effect at the state level. Normalizing the wages of women with a high school degree to 1, we have the following estimates of wages:

	High-School	Some College	College+
Men	1.4653	1.893	2.744
Women	1	1.2931	1.847

Table 2: Wages, CPS 2003 cohort

Finally, we used the PSID data to calculate the fertility rate of different types of households. The sample is defined as the households in 2005 whose head or partner of the head is an appropriately-aged child of the original sample of the PSID. Couples are supposed to have between 0 and 4 children, while singles may have between 0 and 2. The estimated proportions for couples of the same educational level are presented in Table 4.

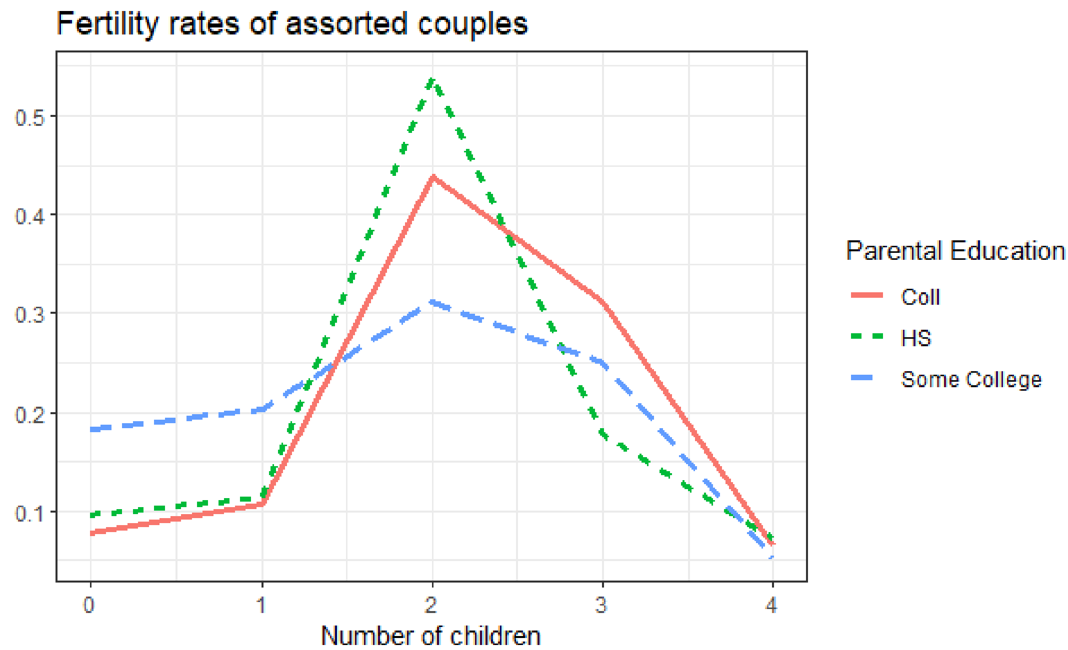

Figure 4: Fertility rates. Data generated with the households formed by children of the original PSID sample.

5 Estimation

Estimation of the model parameters is done in a two-stage process, mimicking the game that the agents play. First, we fix the measure of men and women at the levels found in the data and estimate the parameters of the utility function to fit the marriage pattern and time use data. Then, imposing that this distribution is invariant and given estimated market returns, we calculate the education cost parameters that minimize the distance of the model education choice probabilities from those in the NLSY data.

Let $\Gamma(H_m, H_f)$ be the total economic value received by a couple. The surplus of a marriage of type (H_m, H_f) is defined by the function:

$$\Phi(H_m, H_f) = \Upsilon(H_m, H_f)\Gamma(H_m, H_f) - U(H_m) - V(H_f) - \xi_{H_m,m} - \xi_{H_f,f}$$

The additional terms $\xi_{H_m,m}$ and $\xi_{H_f,f}$ represent non-economic components of the utility of singles, which help to fit the model by capturing components of the utility of singlehood that are common among individuals. To increase the sorting, we multiply the surplus matrix by a multiplicative factor $\Upsilon(H_m, H_f)$ equal to 1 in nondiagonal terms and a positive value on the diagonal. Without these parameters, the surplus estimates are not supermodular, which is incompatible with the assumptions required in the Choo and Siow model for positive sorting.

Moments used are marital patterns, the share of singles in each education level, and the hours of market work and childcare, conditional on the number of children. To account for the heterogeneous patterns of childcare between households, the α is specific to each number of children in the household, the number of parents, and the gender in case they are single.

Given the estimates of the second stage, we include the mean utility of each education level in equation 2. The cost function is linear in education:

$$c_{H,g}(Q) = \delta_{0;H,g} + \delta_{1;H,g}\delta_{nsibs,g}Q$$

Where $\delta_{nsibs,g}$ is a gender-specific multiplicative factor of Q dependent on the number of children of a household. If this factor is smaller for larger numbers of siblings, then childcare is in some way a rival good that is split between children. If it is larger, then it might be the case that children have complementarities that enhance the quality of childcare provided by their parents.

The parameters of the cost function are estimated by minimizing the distance of the choice probabilities of the model from the proportions of the NLSY sample, conditional on the resulting measure of children matching that of the previous generation.

6 Results

In Table 3, we present the predicted childcare and work hours of the model, along with the data moments that were used. The model replicates well the measures of labor supply of the data, with a few exceptions. Child care hours are decreasing in human capital for men and

Table 3: Calibration of the second stage

This table contains the main moments used to calibrate the parameters of the utility functions of the model. Each number represents the expected weekly hours employed in a given activity for some group. The data used comes from the ATUS. Rows represent the education level of husbands, and columns of wives.

Model moments

	Labor M			Labor F		
	HS	SC	CP	HS	SC	CP
HS	43.93	44.84	45.14	24.67	23.75	24.55
SC	45.26	41.92	39.40	22.02	27.55	29.56
CP	45.59	43.37	42.47	17.69	27.76	27.11
	Childcare M			Childcare F		
HS	3.27	3.56	3.81	7.88	8.58	9.17
SC	2.91	3.15	4.12	7.01	7.58	9.92
CP	2.53	3.28	3.57	6.08	7.89	8.59

Data moments

Labor M			Labor F		
41.52	40.49	41.95	21.96	27.79	30.36
46.21	41.13	46.61	24.61	26.45	23.43
41.93	52.97	45.33	25.23	22.27	23.09
Childcare M			Childcare F		
3.11	3.92	5.17	7.71	6.95	9.85
3.04	4.21	7.31	6.56	6.13	10.40
4.02	4.58	7.45	10.04	10.71	12.44

Table 4: Marriage rates and early education patterns

This table contains the marriages generated by the model and those of the data. Each number represents the share of households, with rows representing education level of men and columns of women. The data used comes from the ACS.

Model marriages			Data marriages		
0.1543	0.0754	0.0748	0.1648	0.0668	0.0268
0.0632	0.0805	0.0674	0.0514	0.0901	0.0484
0.0663	0.0734	0.1784	0.0236	0.0601	0.2011

Average early education		
0.0797	0.0868	0.0928
0.0916	0.0990	0.1296
0.1153	0.1495	0.1628

increasing for women. Given the optimal childcare choices shown in Equations 1 and 2, the sign of the change between education levels is entirely dependent on whether the relative increase in income compensates the increase in the price of childcare. As the estimated wages of men vary more steeply than those of women, the resulting estimates behave as expected. However, the expected childcare production matrix for all types of couple is supermodular (as shown in Table 4), which represents the complementarities that we aim to reproduce.

The labor supplies found are a result of the marriage market. In Table 4, we compare the measures of couples in the data and those generated by the surplus matrix of the model. The model generates accurate levels of sorting, though it does overshoot the number of couples where a spouse has college and the other only high school.

With the results of the marriage market, we can then estimate the education cost parameters that rationalize the education choices of individuals. Parameter estimates are included in Table 5. Higher levels of early education make college comparatively less expensive for both genders, as one would expect. The results of the parameter $\delta_{nsibs,g}$ are clear: children with one or no siblings have a boost in their early education, while those with two or more siblings are penalized, indicating that childcare is not a public good for children.

The share of individuals with only high school in the NLSY is significantly greater than the one calculated in the ACS data, which could imply an upward bias for the probability of

$\delta_{0;H,M}$	$\{-0.2127, 2.2864, 4.5849\}$
$\delta_{1;H,M}$	$\{5.4831, -2.3339, -7.1743\}$
$\delta_{0;H,W}$	$\{-3.6901, -1.2614, -0.5043\}$
$\delta_{1;H,W}$	$\{5.3021, -7.4287, -9.6348\}$
$\delta_{nsibs,boys}$	$\{4.0765, 4.0765, 1.2324, 0.5474\}$
$\delta_{nsibs,girls}$	$\{3.8699, 3.8699, 1.2015, 0.6384\}$

Table 5: Cost function estimates

Income statistics	
Gini index of households	0.2141
Gini index of individuals	0.2801
Correlation of income of boys and their fathers	0.0659
Correlation of income of girls and their mothers	0.2041

Table 6: Model estimates of selected income statistics, imposing steady-state

choosing high school. However, the parents of the NLSY sample also have a larger share of high school education, respectively, than the earlier generations of the ACS; therefore, it may be possible that the NLSY only overrepresents such households. As the data used for estimation costs are conditional on parental education (and therefore, it is not simply the education shares), the moments are valid as long as the conditional choice probabilities of the NLSY sample are representative of the overall population.

The income statistics implied by the model are included in Table 6. The Gini index for households of a given generation is quite low, a result possibly explained by the small-range human capital values, which limits the size of the tails of the income distribution. Furthermore, the model equilibrium generates unique positive values of the labor supply for each type of household, which also contributes to reducing the variance of income. The index for individuals is higher, which reflects the high amount of inequality across genders. To measure income persistence, we calculate the correlation of income of individuals of the same sex in families.

6.1 Comparative statics

To understand the effect that sorting may have on persistence, we begin with a mechanical exploration. Suppose that the economy faces a single-period shock to the surplus parameter Υ. Individual decisions of each household remain unchanged, but the utility of being part of a sorted household is modified, which generates a new measure of all households in equilibrium. Although the shock dissipates, the distribution of children changes, which in turn affects the next marriage market, and so forth.

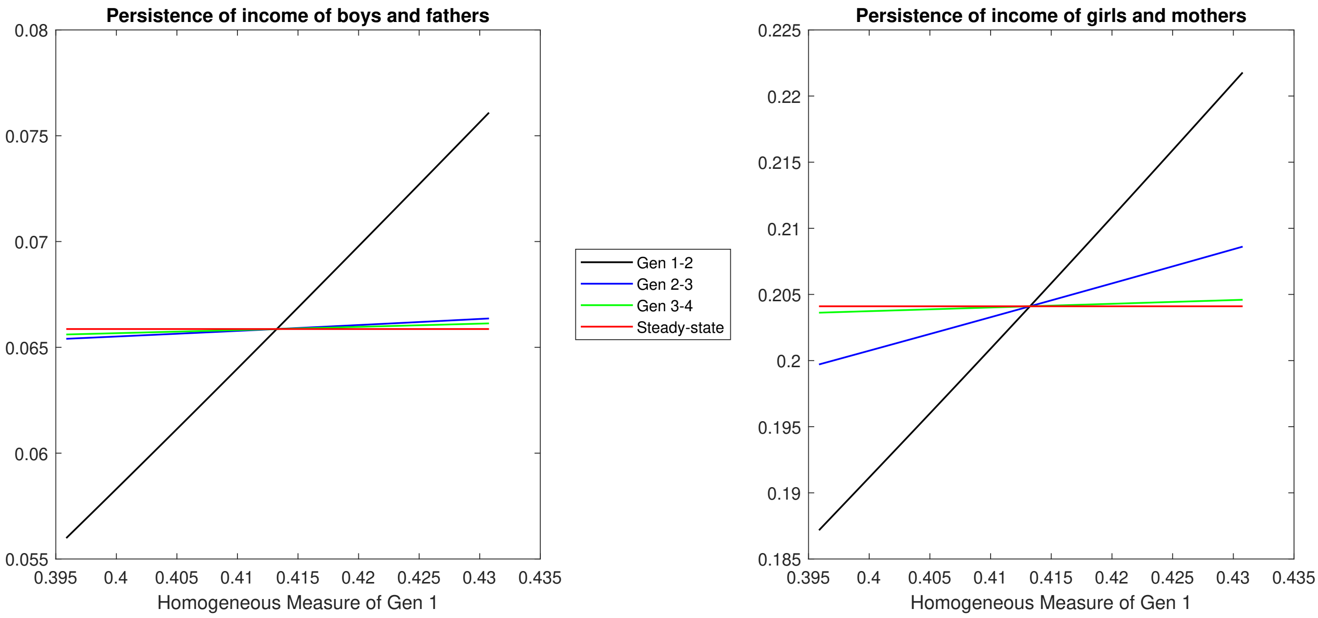

Figure 5: Persistence measures as a function of shocked sorting parameters $\Upsilon(H_m, H_f)$. The horizontal axis represents the total measure of households formed by couples with matching education levels. Each line is the persistence of two consequent generations, with the generation coded as "Gen 1" participating in the altered marriage market.

In Figure 5, we present how the income persistence measures change with the magnitude of the sorting shock. On the horizontal axis, we include the total measure of households whose parents have the same education level, an increasing transformation of this shock. Each line represents the income persistence between two consecutive generations, where "Gen 1" is the generation that matches with the shock present. As we can see, the persistence of both genders is increasing in the overall sorting level of the economy, and the effect takes a few generations to vanish. Evaluating the elasticity of persistence to the total measure of homogeneous couples, we find that an increase of 1% in the size of this population, as a result of an increase in the incentive to sort, generates an increase of 3.6% in the persistence of income of men, and 2% of women.

7 Counterfactuals

Suppose now that we shift the hours that married women spend in housework, an exogenous variable in the model; This can be interpreted as a technological change in home production. The resulting income statistics in the first generation affected by this shock are shown in the blue full line in Figure 6, and the steady-state result is the dotted black line. While personal income inequality decreases as women have more time available to work, household-level inequality

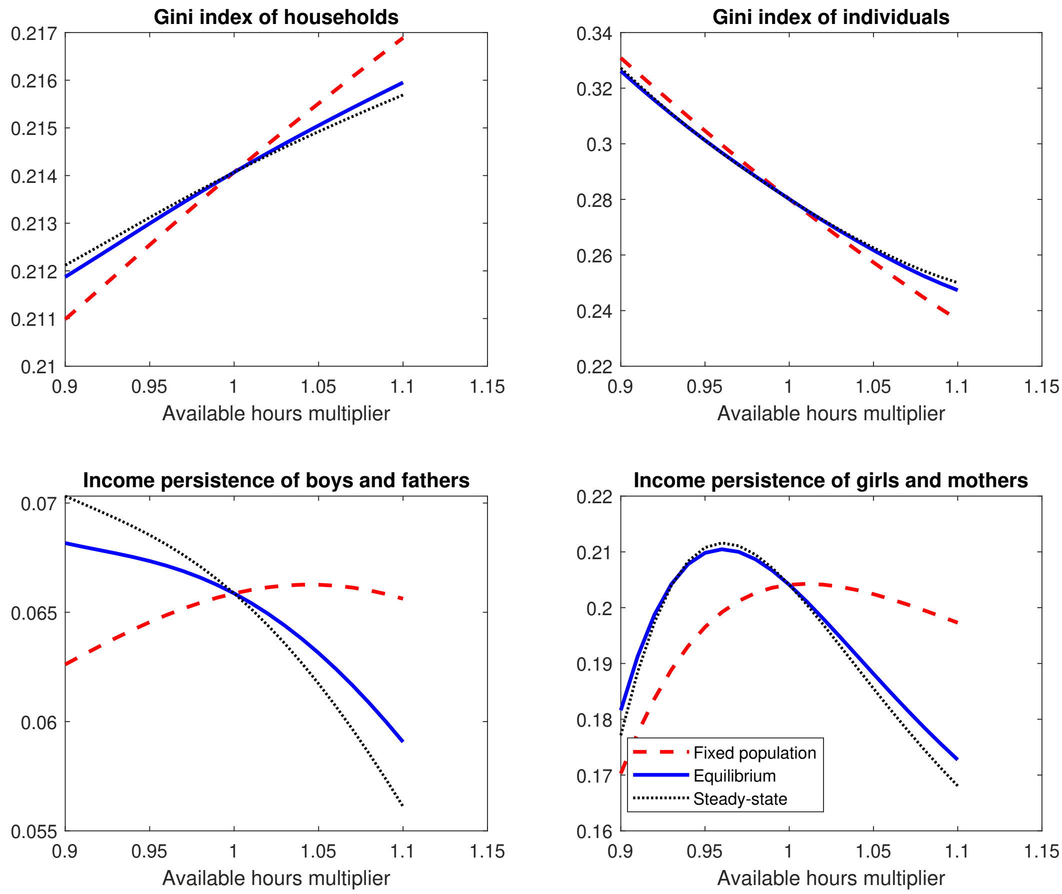

Figure 6: Effects of a shift in available hours for women on income inequality. The dashed line represents an economy with a marriage market formed with the new available hours profile, but a fixed population measure with respect to education level. The bold line shows the results with endogenous population, and the dotted line the steady-state values.

increases. This effect is modulated by the shift in marriage proportions, favoring very high-income households.

The decisions of households will change with the shock, which in turn will modify the educational decisions of each person. If this reduction were to occur instead during the final stage of the game, with individuals having already chosen their education, the distribution of households formed would be generally affected. By contrasting these two situations, we can determine the strength that the endogenous education channel has in the evaluation of this type of counterfactual exercises.

We deploy this variation of the model in the dashed red line in Figure 6. The education level of the population is fixed at the level of the previous steady-state, and they participate in the marriage market with the new parameter levels. The comparison between the two models corresponds to shifting women's housework profile before the two-stage game versus after the first stage.

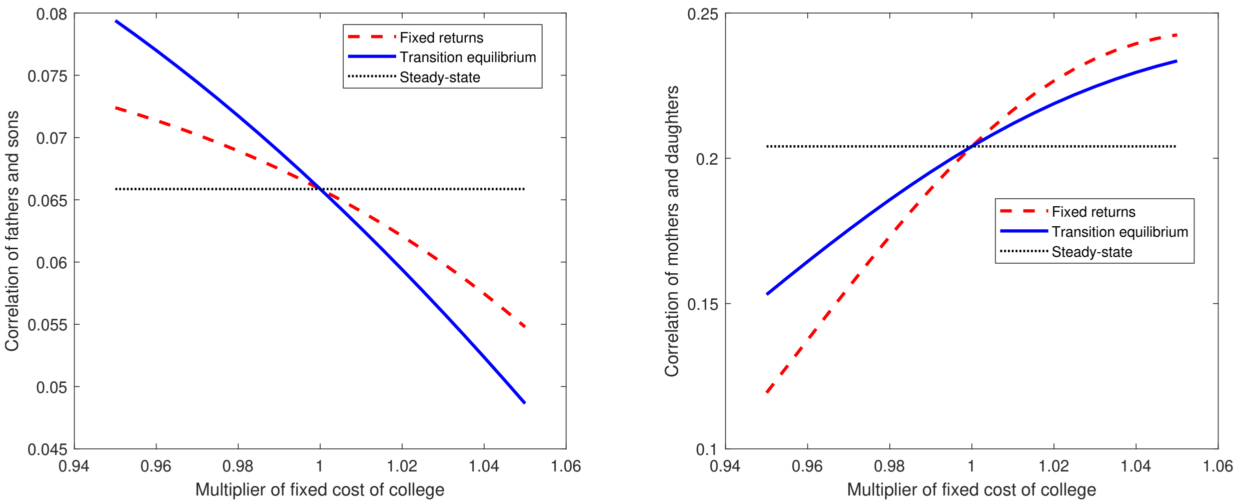

Figure 7: Effect of changes in returns to marriage on income persistence. The parameter that is perturbed is the fixed cost of college for both genders. The bold line is calculated with parents in the original steady-state, and children that choose education as if the average utility of each choice were the same faced by their parents. The dashed line is the standard transition, with agents foreseeing the changes in the marriage market returns. The dotted line is the steady-state.

In both Gini indices, the variation with a fixed population is larger than with adaptation in education decisions. This pattern shows that a model without endogenous choice would overshoot its prediction of effects of the shock, and that the adaptation is towards less heterogeneity. As shown in Appendix B (Figure 8), the full equilibrium is characterized by more sorted college-educated couples and less sorted high school couples, while the partial one presents a general increase in sorting.

In the case of income persistence, there are clear differences in the results between genders. In the equilibrium, almost all households have a decreasing covariance of income of men, while the pattern is less clear with a fixed population, driving the discrepancies seen in the left-down plot. In the case of women, a significant proportion of the effect comes from women in households where the man has a high school degree (see Figures 9 and 10). In this case, the covariance decreases for women that also have only high school, while it increases for better-educated women, as a function of the increase in available hours.

As a last exercise, we evaluate the effects that an increase in the cost of college for men may have. The shift is a value that multiplies the fixed cost of college. This counterfactual is gendered as a way to demonstrate the effect that changes that fundamentals of one gender may have on statistics of the other, as a result of marriage decisions. As shown in Figure 7, persistence of income of men is decreasing in the cost of college for men, while it is increasing for women. Interestingly, the effect has a larger magnitude for women, even though their own costs are unaffected.

To highlight the effect of marriage market forces, we will then decompose these results. Suppose that, starting from the steady-state, there is a transitory shock in the fixed cost of college for men, but each of them makes their education choice without considering that this shift affects all individuals. This is equivalent to choosing as if the expected utility of each level of education in the marriage market remained constant. The income persistence calculated in this generation (where the parents matched in the steady-state model) will change, and we can compare its value to those with correct foresight. Note that this comparison is between two transitory states, and so we also include the steady-state result.

The result in Figure 7 contains interesting dichotomies. The change in persistence is underestimated in the fixed-returns economy for men and overestimated for women. Analyzing the marriage patterns of individuals (presented in Appendix B, Figure 11), we see that the increased cost of college heavily reduces the number of men with college, favoring all other types of couples, in particular those where the woman has a higher level of education.

8 Conclusion

As the marriage patterns of recent decades have shown, individuals have strong tendencies to marry in an assortative fashion. Under the assumptions of friction-less matching, this is interpreted as the presence of complementarities in the utility of couples (in this case, with respect to education). We present a model that generates complementarities through parental investment with children, which attempts to replicate the educational outcomes of children born in the 1960s US while controlling for marriage market forces.

By directly affecting the incentives to sort, we find that small changes in the measure of sorted couples have large effects on the persistence of income of families. These effects cannot be captured in a standard model with dynastic families.

Furthermore, we show that when individuals do not anticipate the effects that shocks have on marriage market returns, the expected effects on income statistics face large errors. This result demonstrates that the exogenous matching of individuals is not a solution, and a market structure for marriages is required.

However, the framework employed in this model is restrictive, due to the hypothesis of transferable utility. Under this structure, we cannot have utility functions as usual in the overlapping generations literature, with parents embedding their children's utility function on their own. It remains to be tested whether this departure generates a significant loss in the predictions of the model.

References

1. S. Rao Aiyagari, Jeremy Greenwood, and Nezih Guner. 2000. "On the State of the Union". *Journal of Political Economy*, 108:2, 213-244

2. Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss. 2017. "Partner Choice, Investment in Children, and the Marital College Premium." *American Economic Review*, 107 (8): 2109-67.

3. Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir. 2018. "The Marriage Market, Labor Supply, and Education Choice." *Journal of Political Economy*, 126(S1), S26-S72.

4. Chiappori, P.-A., Costa-Dias, M., Crossman, S., and Meghir, C. 2020. "Changes in Assortative Matching and Inequality in Income: Evidence for the UK." *Fiscal Studies*, 41, 39-63.

5. Choo, Eugene, and Aloysius Siow. 2006. "Who Marries Whom and Why." *Journal of Political Economy*, 114(1), 175–201.

6. Greenwood, J., Guner, N. and Knowles, J.A. (2003), More on Marriage, Fertility, and the Distribution of Income. International Economic Review, 44: 827-862.

7. Greenwood, Jeremy, et al. 2016. "Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment, and Married Female Labor-Force Participation." *American Economic Journal: Macroeconomics*, 8(1), 1–41.

8. Fernández, Raquel, and Richard Rogerson. 2001. "Sorting and Long-Run Inequality." *The Quarterly Journal of Economics*, 116(4), 1305–1341.

9. Fernández, Raquel, Guner, Nezih, John Knowles, Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality, The Quarterly Journal of Economics , Volume 120, Issue 1, February 2005, Pages 273–344

10. Pilossoph, Laura, and Lin Wee Shu. 2023. "Assortative Matching and Income Inequality: A Structural Approach."

11. Restuccia, Diego, and Carlos Urrutia. 2004. "Intergenerational Persistence of Earnings: The Role of Early and College Education." *American Economic Review*, 94(5), 1354-1378.

Appendix A

Existence of the invariant distribution

Under homogeneous fertility, we can treat each individual of a generation as a single point in the sequence of their family overtime, as if parents pass on to their children. The probability that men of state $\mathbf{x} = (H_m, H_f)$ transition to another $\mathbf{x}^{'} = (H'_m, H^{'}_f)$ (as children) is the probability of their choice to marry that type of individual given their educational status, multiplied by the probability of choosing that education given the returns to education. All state variables matter to individual decision through the education cost functions; therefore, we can consider this a transition from Q to Q'. Note that all the probabilities can be thought of as ex ante to the preference shocks.

$$P(\mathbf{x}^{'}|\mathbf{x}) = P(H^{'}_f|H^{'}_m)P(H^{'}_m|\mathbf{x})$$

These probabilities are divided into 4, in the case of couples and 2, in the case of singles, to account for the leisure preference shocks. This has numerical implications but is not important for the derivations of this section; therefore, we shall omit this additional state for now.

Let us decompose these terms. Given the equilibrium of the marriage market, the first term is the fraction of marriages of this type in the next period $\mu_{H'_m H'_f}$, conditional on the measure of men with this level of education $\mu^{'}_H$.

$$P(H^{'}_f|H^{'}_m) = \frac{\mu_{H'_m H'_f}}{\mu_{H^{'}_m}}$$

Now, the probability of choosing an educational level is the probability of it being the optimal choice. Given GEV type-I shocks, this has the standard logit form. Let $\bar{U}(H)'$ be the mean utility of a choice in the equilibrium of the next period, and $c(H'|x)$ be the cost function of an education choice. This function is dependent on state variables through the choice of childcare expenditure of couples, estimated for each type of marriage.

$$P(H'|\mathbf{x}) = \frac{exp(\bar{U}(H')' - c(H'|\mathbf{x}))}{\sum_{H''} exp(\bar{U}(H'')' - c(H''|\mathbf{x}))}$$

These values are found in equilibrium, which can be found as a result of the following system, where d is the distribution of $\mathbf{x}$, with equivalent equations omitted for brevity.

$$
\begin{aligned}
m_1 - m_1 exp(-\bar{U}(1)) &= \sum_i \sqrt{m_1 exp(-\bar{U}(1)) w_i exp(-\bar{V}(i))} exp(\Phi_{1i}/2) \\
w_1 - w_1 exp(-\bar{V}(1)) &= \sum_j \sqrt{m_j exp(-\bar{U}(j)) w_1 exp(-\bar{V}(1))} exp(\Phi_{j1}/2) \\
m_1 &= \left(\frac{exp(\bar{U}(1) - c(1|d))}{\sum_H exp(\bar{U}(x') - c(x'|d))} \right)' d \\
w_1 &= \left(\frac{exp(\bar{V}(1) - c(1|d))}{\sum_H exp(\bar{V}(x') - c(x'|d))} \right)' d
\end{aligned}
$$

The jacobian matrix of this system with respect to $\bar{U}$, $\bar{V}$, m, and w, can be expressed as

$$
J = \begin{bmatrix} D_u & B \\ C & I_6 \end{bmatrix}
$$

Where I_6 is a 6×6 identity matrix and D_u is a diagonal matrix whose terms are:

$$
\begin{aligned}
&m_i exp(-\bar{U}(i)) + \frac{1}{2} \sum_j \sqrt{m_i exp(-\bar{U}(i)) w_j exp(-\bar{V}(j))} exp(\Phi(ij)/2) \quad \text{for } i \in 1:3 \\
&w_i exp(-\bar{V}(i)) + \frac{1}{2} \sum_j \sqrt{m_j exp(-\bar{U}(j)) w_i exp(-\bar{V}(i))} exp(\Phi(ji)/2) \quad \text{for } i \in 4:6
\end{aligned}
$$

The equations that define the measures are linear combinations of $\mathbf{d}$ with strictly positive terms, therefore there are no types with 0 measure. This implies that the diagonal terms of the matrix D_u are always positive, and it is invertible. As J is a block matrix, it must be invertible, which implies that we can apply the implicit function theorem, providing continuity of the solutions of the system with respect to $\mathbf{d}$.

The result above implies that the right term of $P(\mathbf{x}^{'}|\mathbf{x})$ is continuous. Given the formulas of Choo and Siow (2006), the left term is a continuous function of the right, and so the entire function $F(d_Q)$ is continuous. As it maps the distributions from the simplex $\mathcal{S}^n$ onto itself, which is compact and convex, we can apply the Brouwer fixed-point theorem, which guarantees the existence of $\mathbf{d}^*$ such that $\mathbf{d}^*_{\mathbf{Q}} = F(\mathbf{d}^*_{\mathbf{Q}})$.

Appendix B

Supporting graphs

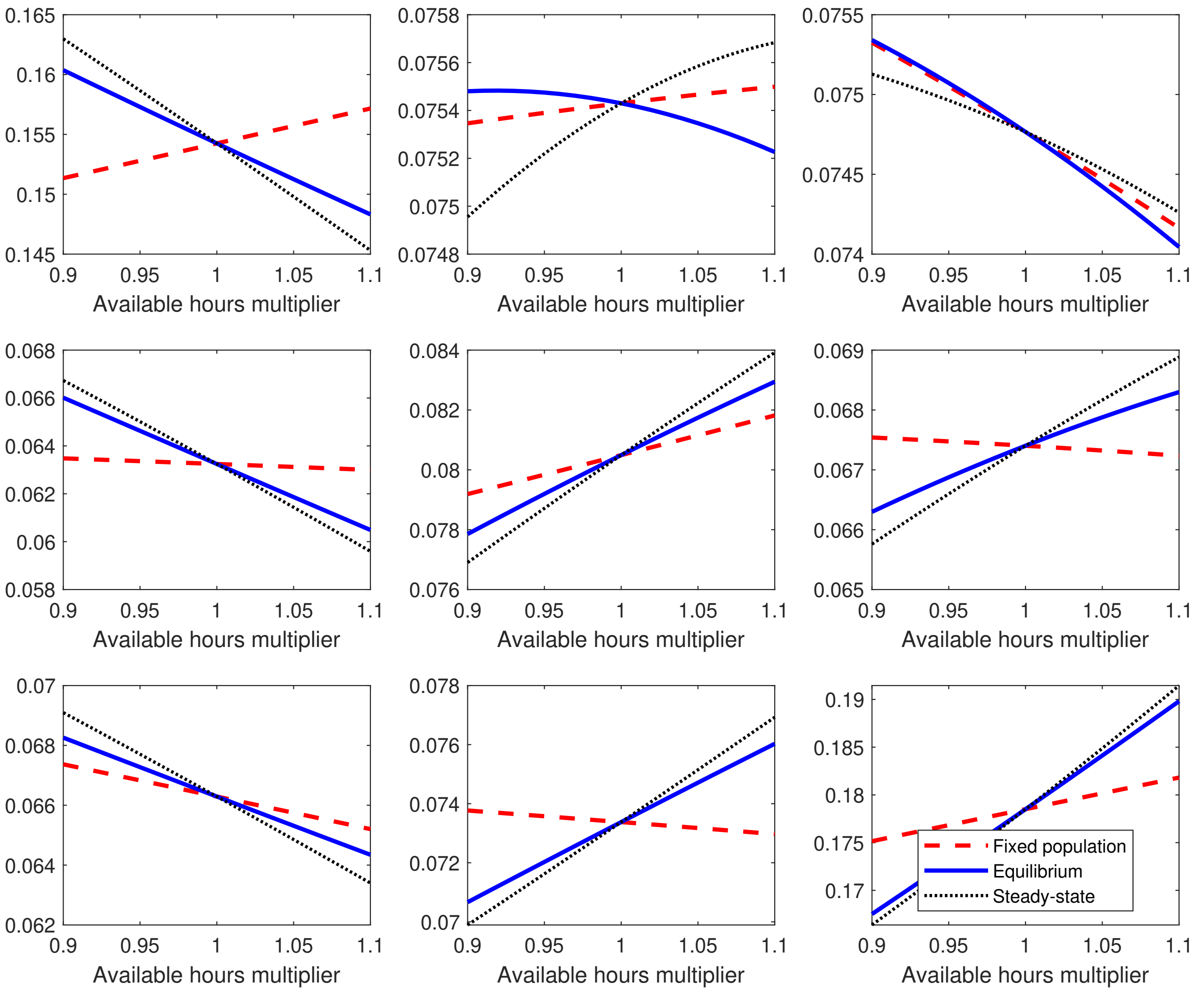

Figure 8: Measure of couple types as a function of the shock to available hours to women. Rows identify the education level of men, from top to bottom: HS, SC, and CP. Columns identify women, from left to right.

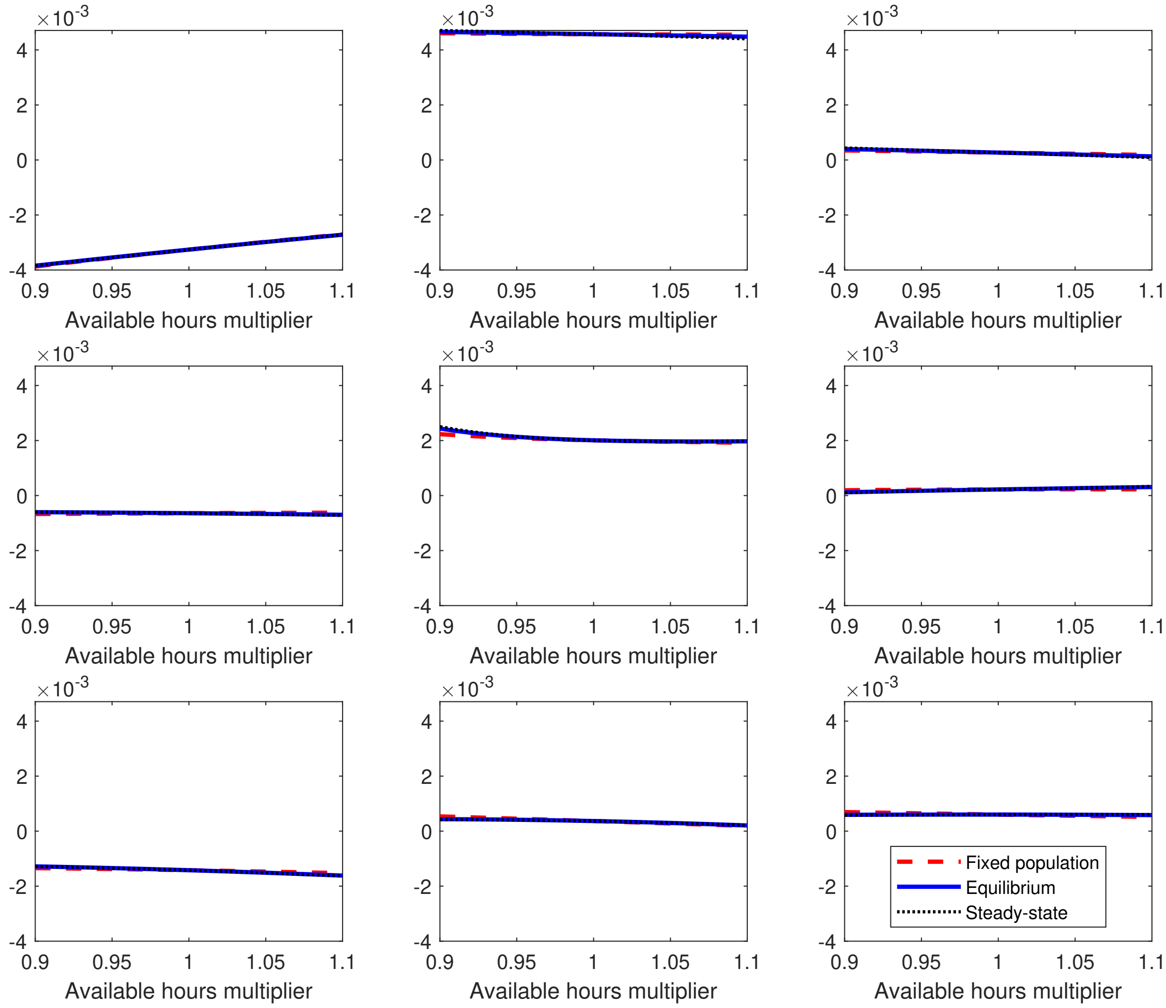

Figure 9: Covariance of income of fathers and boys, conditional on household type of the father, scaled by the measure of such households. Single parents are omitted. Rows identify the education level of men, from top to bottom: HS, SC, and CP. Columns identify women, from left to right.

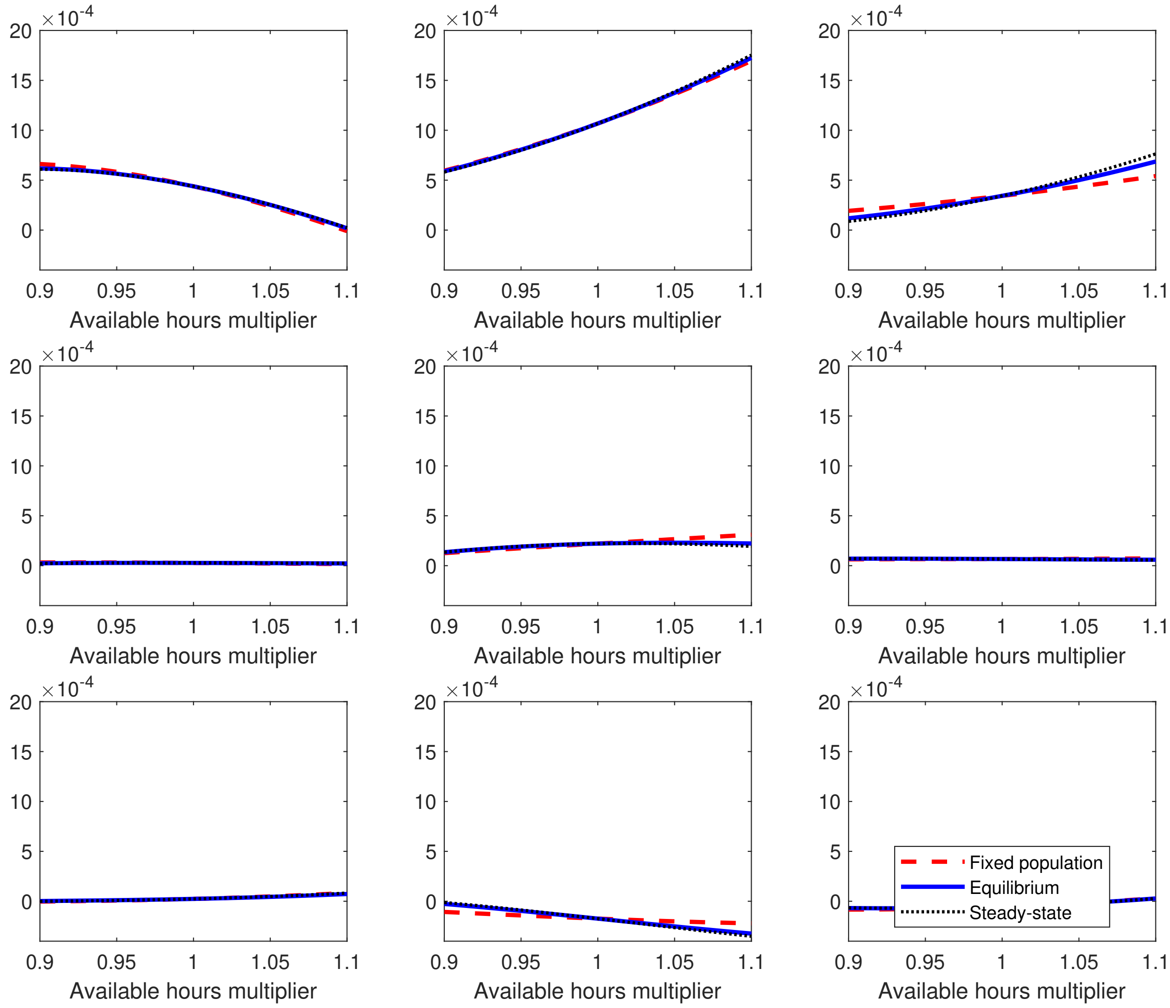

Figure 10: Covariance of income of mothers and daughters, conditional on household type of the mother, scaled by the measure of such households. Single parents are omitted. Rows identify the education level of men, from top to bottom: HS, SC, and CP. Columns identify women, from left to right.

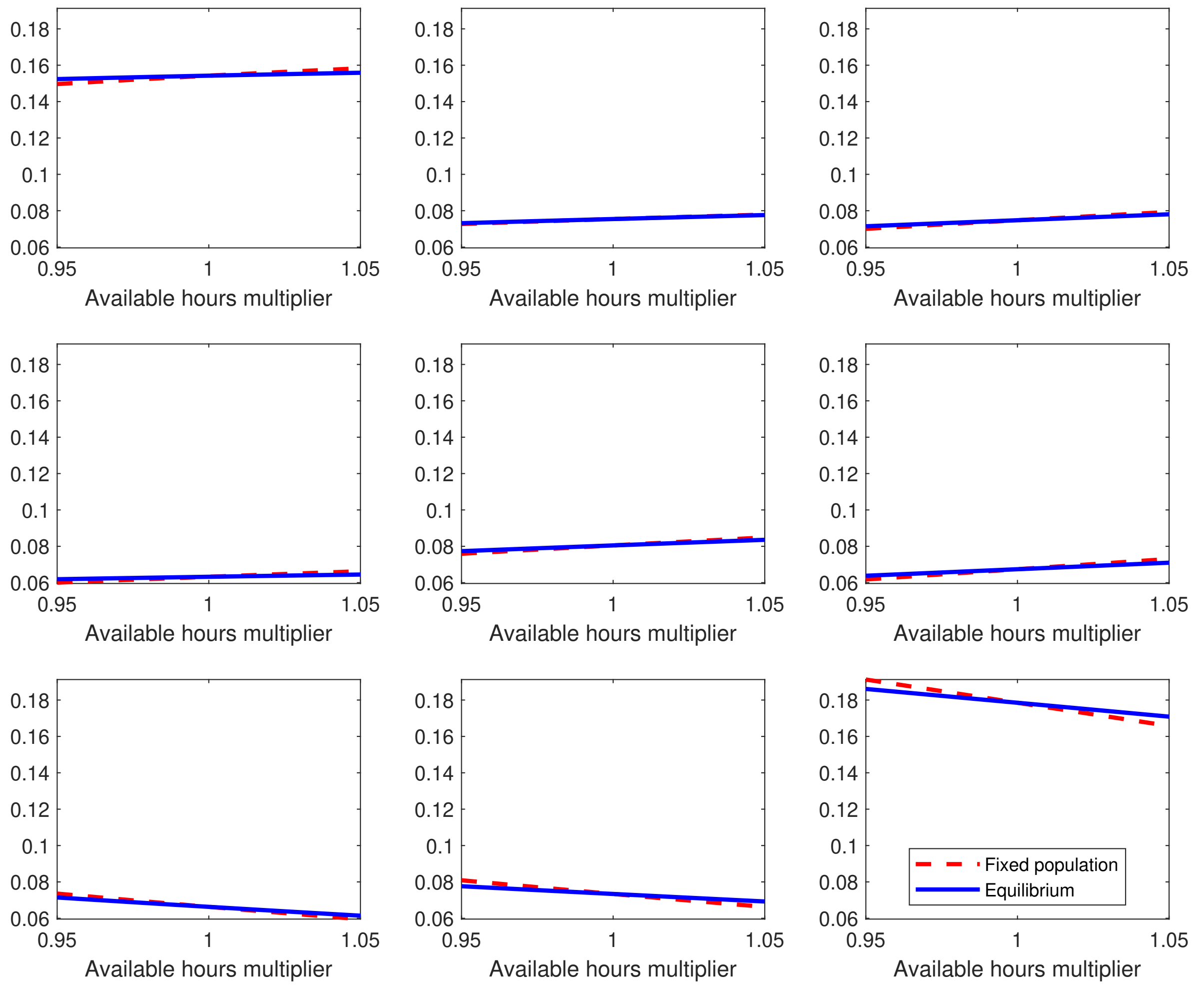

Figure 11: Measure of couple types as a function of the shock to college costs of men. Rows identify the education level of men, from top to bottom: HS, SC, and CP. Columns identify women, from left to right.

Appendix C

Parameter estimates

β	0.2937
ξ	0.0260
λ	0.0014
σ	-12.4167
α of couples, per number of children (0-4)	$\{0.0173, 0.0558, 0.0589, 0.1152, 0.1992\}$
α of single men, per number of children (0-2)	$\{0.6170, 1.7452, 1.4632\}$
α of single women, per number of children (0-2)	$\{1.0539, 1.9980, 3.7413\}$
Bonus utility of single men per education level	$\{-0.1931, -0.3506, -0.7715\}$
Bonus utility of single women per education level	$\{-0.0772, -0.0922, -0.5872\}$
Multiplier of surplus of sorted couples	$\{2.6180, 1.2216, 1.9548\}$