Segregation and Sorting of U.S. Households: Who Marries Whom and Where?*

Davide Alonzo — Université de Montréal

Nezih Guner — CEMFI

Claudio Luccioletti — CEMFI

November 22, 2023

Preliminary Draft: Latest Version

Abstract

We use a spatial heterogeneous agent model of family formation to study the role of geography in shaping trends in marriage and inequality in the US. The United States has undergone dramatic shifts in household and family structures since the 1980s. Married female labor force participation increased significantly, marriage has declined, and positive assortative mating has risen. Along with a rising skill premium and a declining gender gap, income inequality among households has also widened. We document that these changes had an important geographic dimension: the decline in marriage was much more significant in smaller cities, and marital sorting declined in smaller cities but increased in larger ones. We interpret these facts within a model where households decide where to live, whether or not to get married, and with whom. The framework allows us to quantify the importance of each channel in shaping trends in marriage and inequality.

Keywords: Inequality; Marriage; Labor Force Participation, Spatial Dispersion, Great Divergence.

JEL Classification: J11; J12; R13; R23.

*Davide Alonzo: davide.alonzo@umontreal.ca, Nezih Guner: nezih.guner@cemfi.es, Claudio Luccioletti: claudio.luccioletti@cemfi.edu.es. We thank seminar participants at CEMFI for their helpful comments. Aitana Simó Bartual and Guillem Tobías Larrauri provided excellent research assistance.

1 Introduction

US households today look very different from how they were decades ago. Rising returns to skills and declining employment in manufacturing increased the earnings dispersion between individuals with and without a college degree. At the same time, the share of the married population declined significantly, and much more so for less-educated individuals. There has also been a rise in assortative mating as couples are much more likely to have the same educational levels. All these forces generated a growing income inequality among households. Those with less education either remain single or match with other less educated individuals, while college graduates are more likely to be married to others like themselves. A countervailing force to these trends has been the growing labor force participation of married females and the declining gender wage gap, which reduced household-level income inequality.

There has also been a geographical dimension to all these changes. Since the 1980s, the share of college-educated individuals has become much more concentrated in bigger cities. Inequality also increased more in bigger cities. While these trends are well known, we document that changes in marriage patterns also differ significantly across cities.

The decline in marriage has been much more significant in smaller cities, and, in particular, the share of non-college graduates who are single in small cities more than doubled between 1980 and 2019. Furthermore, the increase was much more pronounced in larger cities. In 1980, assortative mating was lower in larger cities than in smaller ones. Today, it is much higher in larger cities. Finally, housing price dispersion across US cities increased, with prices rising much more rapidly in bigger cities.

In this paper, we use a spatial heterogeneous agent model of family formation to study the role of geography in shaping trends in marriage and inequality in the US. In the model, individuals who differ by innate ability first choose whether or not to obtain a college degree and where to live. Locations (cities) differ in their amenities, which are exogenous. They also differ in the wages they offer workers with different skills and genders, which reflects both exogenous differences in technology and the endogenous distribution of population across space. Once in a city, singles meet at a marriage market and form married couple households. Some marriages end in divorce, and individuals can remarry. Households, married or single, decide how much market goods and housing to consume. Married couples also make labor force participation decisions for both partners. Finally, housing prices are determined by city-specific housing supply elasticities, which are exogenous in the model, and endogenous housing demand.

We estimate the model parameters by targeting several aggregate and cross-sectional moments from the 2019 US data. We match wages by skills and gender as well as housing prices across cities. Furthermore, the model can generate the observed population distribution by skill and marital status across US cities. The degree of marital sorting and labor force participation by city size in the model is also consistent with what we observe in the data. At the aggregate level, the share of the skilled population, the extent of earnings inequality, and the share of housing expenditure by household income are targeted.

We next use the model economy to study how household and location decisions changed in the US between 1980 and 2019. To this end, we conduct two counterfactual experiments. First, we change the technology parameters so that the model economy matches the observed level of wages, skill premium, and gender wage gap in 1980. Second, we change the elasticity of the housing supply in each city and set them to their 1980 values. The housing supply elasticities in 1980 were much higher, in particular in smaller cities.

Changes in wage structure between 1980 and 2019 generate a much larger share of individuals with a college degree and a significant increase in married female labor force participation. In the model, as in the data, the share of college graduates increases more in bigger cities. The share of the population who choose to get married declines. The assortative mating increases significantly, and much more so in bigger cities, as we observed in the data. These changes also affect the housing market. House prices increase everywhere, but the increase is much larger in bigger cities. This generates a greater dispersion of house prices across cities, which is also consistent with what has happened in the US.

The decline in the housing supply elasticities between 1980 and 2019 generate a significant increase in house prices. In the model, this increase affects household decisions. On the one hand, with higher housing prices, labor force participation of married females increases as households try to cover a higher expenditure on housing. The increase in married female labor force participation is larger in bigger cities with higher housing prices. At the same time, higher housing prices in 2019 also generate more assortative mating, particularly in bigger cities. With higher housing prices, skilled individuals are more likely to form households with two skilled wages.

The paper builds on two strands of literature. Our analysis is related to a large number of papers that study changes in US marriage patterns, building and estimating equilibrium matching models of household formation and dissolution. Greenwood et al. (2016), Goussé, Jacquemet and Robin (2017), and Blasutto (2023) are examples in this literature. We also build on papers that study spatial changes in the US labor markets, such as Diamond (2016), Eckert, Ganapati and Walsh (2022), and Giannone (2022). A set of recent papers shares our focus on how household decisions interact with the location choices of households. Moreno-Maldonado (2023) studies labor force participation of married women with children across cities, focusing on why it is lower in bigger cities. Moreno-Maldonado and Santamaria (2022) study how delayed fertility has contributed to the gentrification of city centers in the US by delaying movements of couples to suburbs. More directly related to the current paper, Alonzo (2022) and Fan and Zou (2022) build models of marriage with location decisions. Alonzo (2022) studies how location-specific returns to occupations affect marriage and location decisions, while Fan and Zou (2022) focus on how joint marriage and location decisions affect the spatial dispersion of economic activity.

2 U.S. Households across Time and Space

Three key trends have been shaping the US society since the 1980s. First, there has been a significant increase in inequality, which has been extensively documented and discussed in the literature – see, among others, Acemoglu and Autor (2011), Heathcote, Perri and Violante (2010), and Heathcote et al. (2023). Whatever measure one looks at, whether it is the skill premium or the share of income received by top percentiles of the income distribution or more broad measure such as the Gini coefficient, the inequality increased significantly between 1980 and 2010 and has been relatively stable since then.

Second, a much smaller fraction of the population is married today than in the past. The decline in marriage has been much more pronounced among the less educated individuals. In 1980, 67% and 71% of females between ages 25 and 54 with and without a college degree were married, respectively. Today, only 62% and 50% of them are. With the decline of the married population, who is married with whom also changed. There has been an increase in assortative mating, as it is more likely for people with similar educational attainment to form marriages – Greenwood et al. (2014), Chiappori, Dias and Meghir (2020).

Finally, there has been a significant decline in the labor force participation of less educated men. In 1980, 92% of men without a college degree between ages 25 and 54 were in the labor force. Today, only 72% of them are in the labor force. Krueger (2017) and Charles, Hurst and Schwartz (2019) analyze the decline in work among less educated men.

These changes, however, did not take place uniformly across the US geography. Since the 1980s, the share of the US population with a college degree and the skill (college) premium have increased significantly. At the same time, as shown in Figure 1, the relation between city size and the share of college-educated population has become increasingly positive (the so-called great divergence, Berry and Glaeser, 2005). In 1980, there was no clear relation between city size and skill premium. Since then, a positive relation between city size and the skill premium has emerged, as shown in Figure 1b.

A positive relation between city size and inequality has also emerged since the 1980s, as already highlighted by Baum-Snow and Pavan (2013) and Baum-Snow, Freedman and Pavan (2018). Figure 2 shows these patterns for the 50-10 and 90-10 earnings ratios.

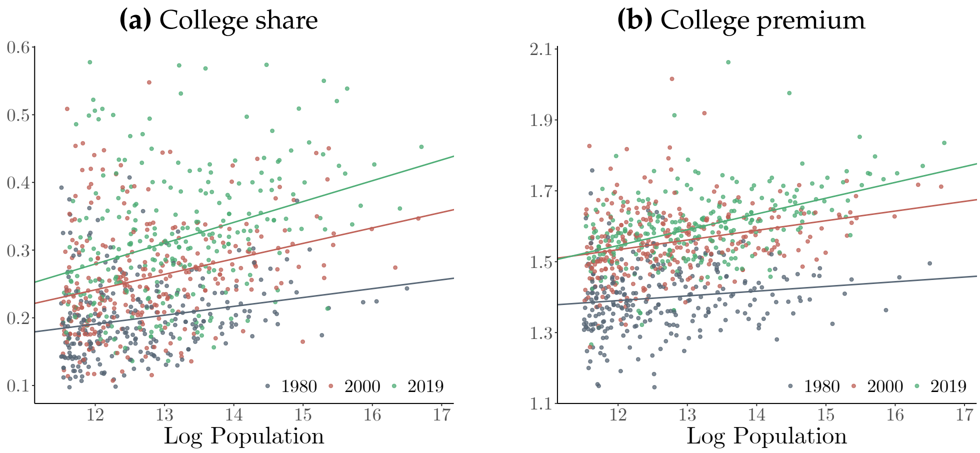


Figure 1: The figures plot MSA outcomes by year. Sample: 25-54 years old. Data: 1980, 2000, ACS 5-years 2015-2019.

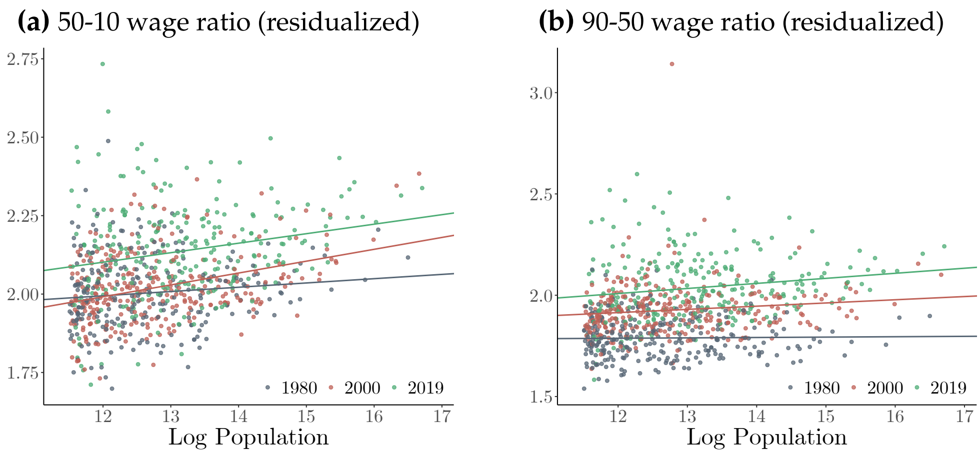
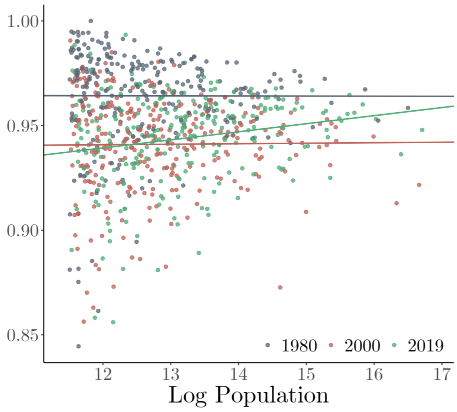


Figure 2: The figures plot MSA outcomes by year. Sample: 25-54 years old, full-time workers employed for more than 48 weeks. Wage is residualized by gender and marital status (interacted), race, education level, and experience (linear and squared). Data: Census 1980, 2000, ACS 5-years 2015-2019.

Changes in employment levels were also not uniform across cities. While the share of men with full-time employment declined since the 1980s, a positive relation between city size and employment emerged in recent decades, mainly due to a decline in jobs in smaller cities (Figure 3). An important factor behind the decrease in overall employment has been the reduction in manufacturing jobs (Figure 4).

In contrast to the decline in employment among men, there has been a significant increase in the share of women who are employed full-time, with no clear relation with city size (Figure 5). Along with higher employment opportunities for women, the gender wage gap has declined since 1980. While the relationship between city size and the gender gap was negative in 1980, it became less negative in 2000 and was flat in 2019.

(a) Share employed males, college



(b) Share employed males, non-college

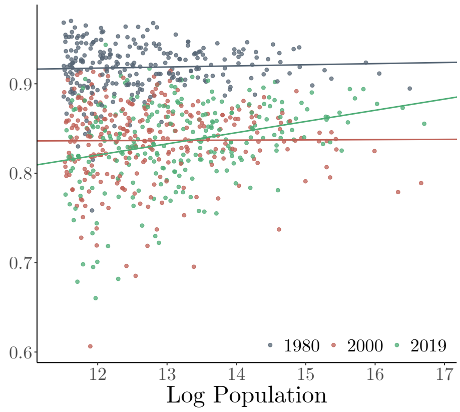
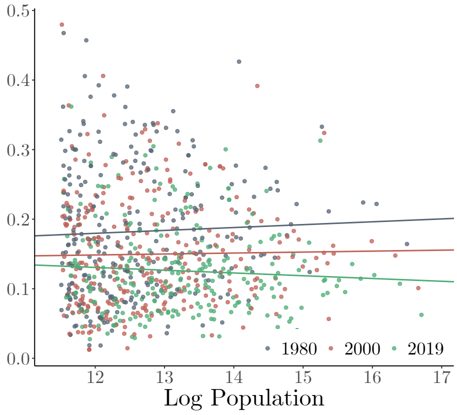


Figure 3: The figures plot MSA outcomes by year. Sample: 25-54 years old. Data: 1980, 2000, ACS 5-years 2015-2019.

(a) Manufacturing employment share, Males college



(b) Manufacturing employment share, Males non-college

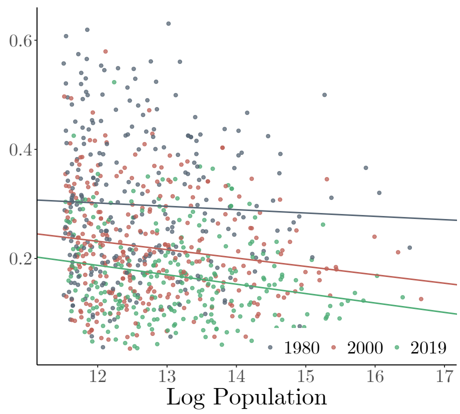
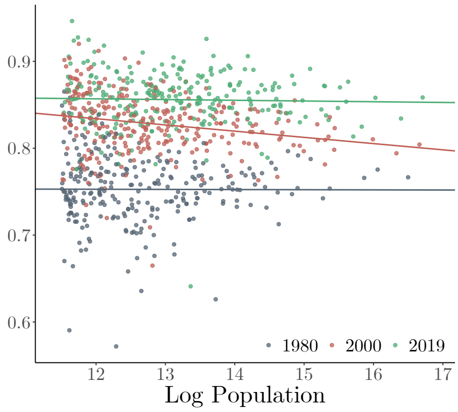


Figure 4: The figures plot MSA outcomes by year. Sample: 25-54 years old. Data: 1980, 2000, ACS 5-years 2015-2019.

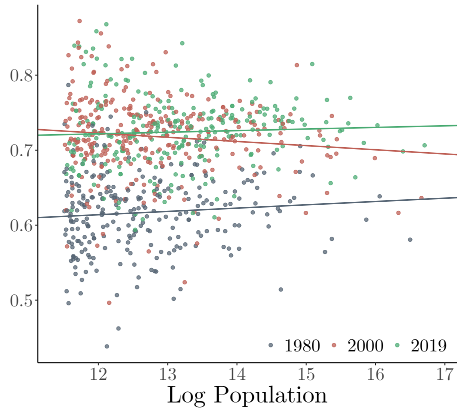
Moreover, housing price dispersion across US cities increased since the 1980s, with prices rising much more rapidly in bigger cities (Van Nieuwerburgh and Weill, 2010). As shown in Figure 6, the positive relationship between prices and city size has become steeper in recent decades.

Between 1980 and 2019, the fraction of the population between ages 25 and 54 who are married declined significantly. But the decline has been more pronounced in smaller cities. As a result, while the relation between city size and marriage was negative in 1980, i.e., the number of singles was higher in bigger cities, it became flat in 2019 (Figures 7a and 7b). The pattern is similar for college and non-college, but the

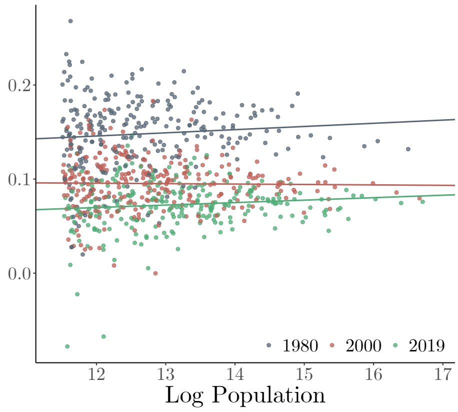
(a) Share employed females, college



(b) Share employed females, non-college



(c) Gender wage gap, college



(d) Gender wage gap, non-college

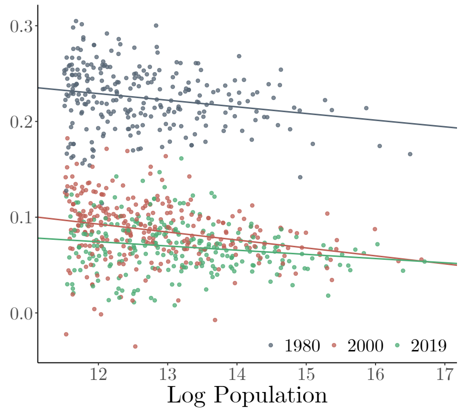


Figure 5: The figures plot MSA outcomes by year. Sample: 25-54 years old. Data: 1980, 2000, ACS 5-years 2015-2019.

(a) Annual rent (thousands)

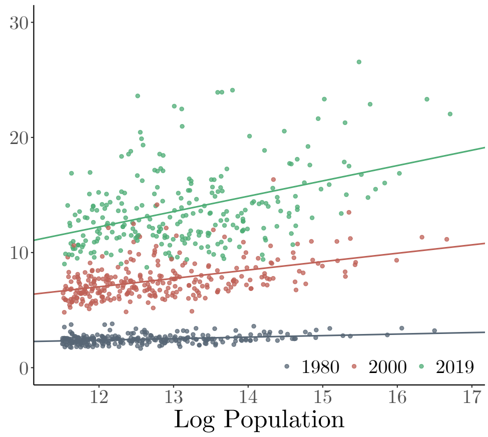


Figure 6: The figure plot MSA outcomes by year. Annual rents are adjusted for dwelling characteristics using a hedonic regression (see Appendix A). Sample: 25-54 years old. Data: 1980, 2000, ACS 5-years 2015-2019.

decline in marriage has been much more significant for non-college. At the same time, the number of divorced people increased, and the rise was more significant in smaller cities. Indeed, the relationship between city size and divorce was positive in 1980 and became negative in 2019. The pattern is similar for college and non-college, but again, the rise in divorce is more significant for non-college (Figures 7c and 7d).

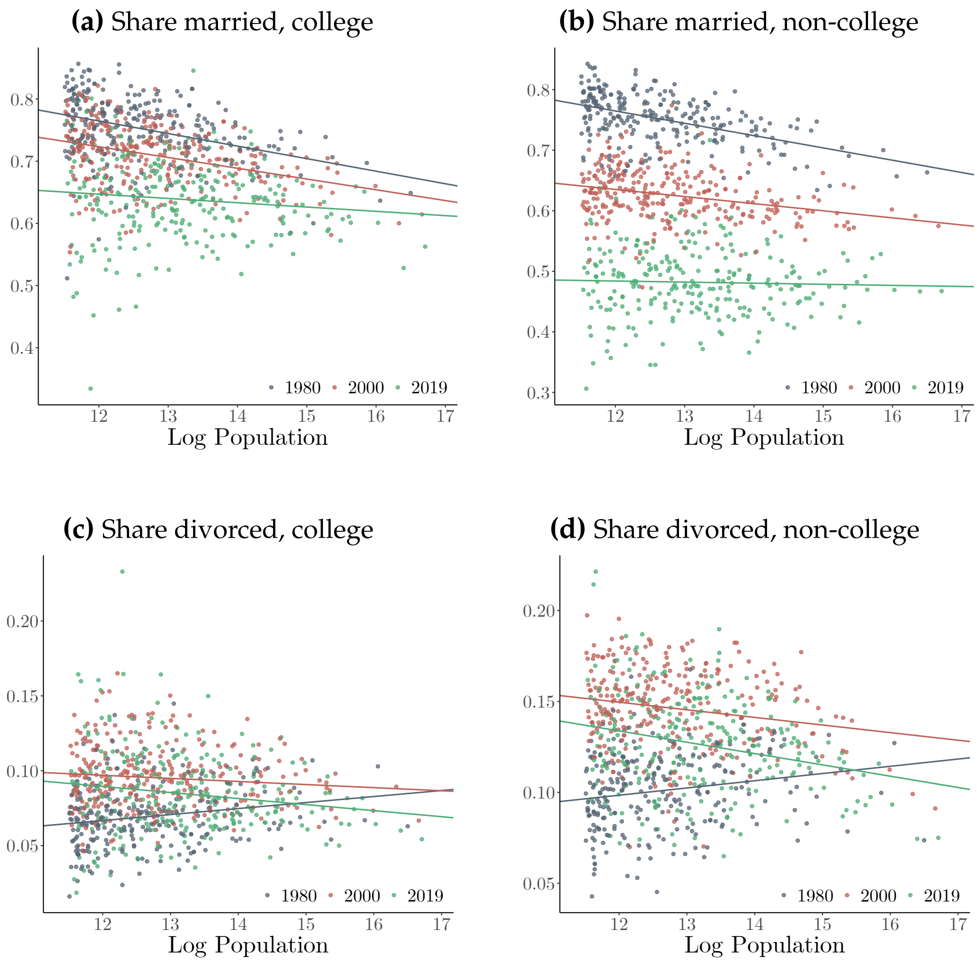


Figure 7: The figures plot MSA outcomes by year. Sample: 25-54 years old. Data: 1980, 2000, ACS 5-years 2015-2019.

Finally, there was no relation between marital sorting (the correlation between the educational attainments of couples) and the city size in 1980. Since then, the relationship has become positive, and there is now a higher degree of assortative mating in larger cities, as marital sorting declined in smaller cities and increased in larger ones (Figure 8). The increase in the number of couples with two college graduates, the so-called power couples, has already been documented by Costa and Kahn (2000) and Compton and Pollak (2007).

(a) Marital sorting by education (correlation)

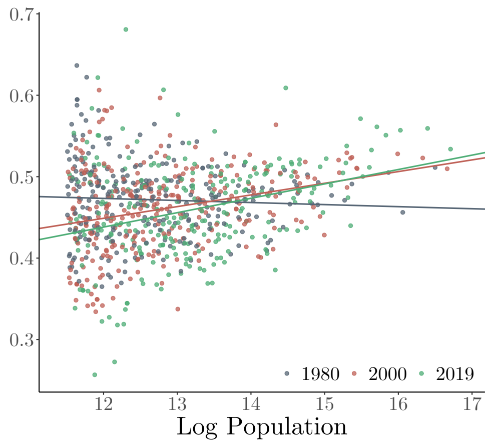


Figure 8: The figure plot MSA outcomes by year. Sample: 25-54 years old. Data: 1980, 2000, ACS 5-years 2015-2019.

3 Model

The economy is populated by infinitely-lived individuals who differ in gender, $g \in \{m, f\}$. Individuals are born with an innate ability level, denoted by a_g. We assume that $\ln a_g \sim N(0, \sigma_a)$. At the start of their lives, individuals decide whether to get a college education, denoted by $e \in \{N, E\}$. Obtaining a college education implies a utility cost that is declining in innate ability. After the education decision, individuals decide where to live. There are J cities, indexed by $j \in \{1, 2, ..., J\}$. Cities differ in the amenities they offer. Wages for individuals of different genders and education also differ across cities. These wages depend on exogenous productivity differences across cities and endogenous education and location decisions. Labor incomes are a product of wages and innate ability levels.

Finally, once individuals choose a city, they enter a local marriage market, meet other singles, and form households. In each period, single and married households decide how much market goods and housing to consume. Married households also decide whether or not their members participate in the labor market, while single households are assumed always to participate. Marriages can also break, and if that happens, divorcees reenter the marriage market. We assume that education and location decisions are made once at the start of life, while marriage and household decisions are made continuously. Each period, individuals can die at an exogenous Poisson arrival rate of π, and are replaced by newborns.

3.1 Preferences

The instantaneous utility function for a single individual in city j is given by

$$u_j^S(c,h) = \alpha \log(c) + (1-\alpha)\log(h - \underline{h}) + \beta B_j,$$

where c and h are the consumption of market goods and housing. There is a minimum housing consumption, denoted by $\underline{h}$, which captures the fact that poorer households spend a larger share of their resources on housing. Finally, B_j represents the amenities that city-j offers.

For a married couple, the instantaneous utility from consumption and housing of living in city j is

$$u_j^M(c,h,l_f,l_m) = \alpha \log(c) + (1-\alpha)\log(h - \underline{h}) + \beta B_j + q_f(1-l_f) + q_m(1-l_m),$$

where l_f and l_m are labor supply decisions, i.e., $l_g \in \{0,1\}$ and q_f and q_m are utility gains from choosing not to work. A married couple also enjoys an additive match quality, denoted by ε^M such that the total instantaneous utility is $u_j^M(c,h,l_f,l_m) + \varepsilon^M$.

3.2 Production

The output in location j is given by

$$Y_j = X_j(\theta_j^N L_{j,N}^\rho + (1-\theta_j^N)L_{j,E}^\rho)^{1/\rho},$$

where X_j is city-specific productivity, θ_j^N is the share of non-college educated workers in production and ρ is the elasticity of substitution between college and non-college workers. The total efficiency units supplied in city j by education group e, $L_{j,e}$, is an endogenous aggregate that depends on how individuals of different ability and education levels are distributed across space. These efficiency units aggregate male and female labor supply, i.e.,

$$L_{j,E} = L_{j,E,m} + \kappa_j^E L_{j,E,f},$$

and

$$L_{j,N} = L_{j,N,m} + \kappa_j^N L_{j,N,f},$$

where κ_j^e is the relative efficiency of females with respect to males, a parameter that can help us to match the observed gender wage gaps.

Labor and product markets are competitive and there is a representative firm in each location. As a result, the wages in each location are given by the following first-order conditions for unskilled

$$w_{j,N,m} = X_j{}^{\rho}\theta_j^N \left(\frac{Y_j}{L_{j,N}}\right)^{1-\rho}, \quad w_{j,N,f} = X_j{}^{\rho}\theta_j^N \kappa_j^N \left(\frac{Y_j}{L_{j,N}}\right)^{1-\rho},$$

and skilled workers

$$w_{j,E,m} = X_j{}^{\rho}(1-\theta_j^N) \left(\frac{Y_j}{L_{j,E}}\right)^{1-\rho}, \quad w_{j,E,f} = X_j{}^{\rho}(1-\theta_j^N)\kappa_j^E \left(\frac{Y_j}{L_{j,E}}\right)^{1-\rho}.$$

In each location, housing supply is given by

$$H_j = \omega_j^0 p_j^{\omega_j^1},$$

where p_j is the price per unit of housing h, ω_j^1 is the elasticity of housing supply, and ω_j^0 is a city-specific scaling parameter that will allow us to exactly reproduce the housing prices observed in the data.

3.3 Marriage Matching

As a result of education and location decisions, and the formation and dissolution of households, in each period there will be a mass of single males and single females in each location. Let μ_{j,e_f,a_f} denote the mass of single females of education level e and innate ability a in city j, and μ_{j,e_m,a_m} be the corresponding number for males. Then, the total number of singles of gender g in city j is given by

$$\mu_{j,g} = \sum_{e_g} \sum_{a_g} \mu_{j,e_g,a_g}.$$

The number of meetings among singles of particular education and ability per unit of time, $\Gamma_j(e_f, a_f, e_m, a_m)$, is determined by a Cobb-Douglas matching function,

$$\Gamma_j(e_f, a_f, e_m, a_m) = \psi(e_m, e_f) \left(\mu_{j,f}^{\beta}\mu_{j,m}^{1-\beta}\right) \frac{\mu_{j,e_f,a_f}}{\mu_{j,f}} \frac{\mu_{j,e_m,a_m}}{\mu_{j,m}},$$

where

$$\psi(e_m, e_f) = \begin{cases} \psi(1+\phi_j), & \text{if } e_m = e_f \\ \psi(1-\phi_j), & \text{otherwise} \end{cases}.$$

The parameter ψ represents the efficiency of the matching function, and ϕ_j captures the fact individuals with similar education levels are more likely to meet each other.

Given the number of matches formed, the probability (Poisson rate) that a single woman of skill e_f an ability a_f matches with a single man of skill e_m an ability a_m is given by

$$\xi_{j,f}(e_f, a_f, e_m, a_m) = \frac{\Gamma_j(e_f, a_f, e_m, a_m)}{\mu^S_{j,e_f,a_f}}.$$

3.4 Taxes and Transfers

Households face a progressive Federal income tax schedule. Following Bénabou (2002), Heathcote, Storesletten and Violante (2014), and others, we use a simple tax function to represent effective average tax rates in the data given by

$$t(\tilde{I}) = 1 - (1 - \tau_0)\tilde{I}^{-\tau_1}, \tag{1}$$

where $\tilde{I}$ is the income of a household relative to the mean income in the economy. The parameter τ_0 defines the "level" of the tax rate, whereas the parameter τ_1 governs the curvature or progressivity of the system. If $\tau_1 = 0,$ then taxes are proportional and given by τ_0. If τ_1 is positive, taxes are progressive and higher-income households face a higher average tax rate.

Households also receive means-tested transfers. Following Guner, Lopez-Daneri and Ventura (2023), we assume that the effective transfer function takes the following form

$$TR(\tilde{I}) = \begin{cases} \gamma_0, \text{ if } \tilde{I} = 0 \\ \exp(\gamma_1)\exp(\gamma_2 \tilde{I})\tilde{I}^{\gamma_3}, \text{ if } \tilde{I} > 0 \end{cases},$$

where $\tilde{I}$ again is household income relative to the mean income. This formulation implies that transfers are positive if household income is zero and accommodates a smooth decline as household income increases. Below, we use function $T(I)$ to denote total tax and transfer payments households pay or receive.

3.5 Household Decisions

We first define the dynamic programming problem of single and married households after they make their location decisions. The structure of the value functions closely follows the continuous-time marriage model by Goussé, Jacquemet and Robin (2017).

3.5.1 Single households

After choosing a city to live in, a single female continuously chooses how much market goods and housing to consume. With some probability (Poisson rate) $\xi_{j,f}(e_m, a_m, e_f, a_f)$, a single female of type (e_f, a_f) is matched with a type-(e_m, a_m) man in the marriage market. Upon matching, the marriage quality parameter $\varepsilon^M \sim N(\mu_{\varepsilon^M}, \sigma_{\varepsilon^M})$ is drawn, and, after observing it, the couple decides whether to marry. Let $V^M_{j,g}(a_m, e_m, a_f, e_f, \varepsilon^M)$

and $V^S_{j,g}(a_g, e_g)$ denote the value of being married and single, respectively. A marriage takes place if

$$V^M_{j,g}(a_m, e_m, a_f, e_f, \varepsilon^M) \geq V^S_{j,g}(a_g, e_g) \text{ for } g \in \{f, m\}, \tag{2}$$

i.e., both parties agree.

Since $V^M_{j,g}$ is strictly increasing in ε^M, this implies a threshold rule for marriage. Let $\bar{\varepsilon}(a_m, e_m, a_f, e_f)$ be this threshold such that a marriage occurs if $\varepsilon^M \geq \bar{\varepsilon}(a_m, e_m, a_f, e_f)$. As a result, upon matching, marriage happens with probability

$$1 - \Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right),$$

where $\Phi(\cdot)$ is the CDF of the normal distribution.

Given a rate of time preference r and Poisson arrival rate for death π, the value function for a single female is given by

$$\begin{aligned}(r+\pi)V^S_{j,f}(a_f, e_f) &= \max_{c,h} u^S_j(c,h) + \sum_{a_m, e_m} \xi_{j,f}(e_m, a_m, e_f, a_f) \\ &\int_{\bar{\varepsilon}(a_m, e_m, a_f, e_m)}^{\infty} \left[V^M_{j,f}(a_m, e_m, a_f, e_f, \tilde{\varepsilon}) - V^S_{j,f}(a_f, e_f)\right] d\Phi\left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right) \\ \text{s.t.} \quad & c + p_j h = w_{j,e_g,g} a_g + T(w_{j,e_g,g} a_g)\end{aligned} \tag{3}$$

Consumption and housing decisions are static, the allocations are independent of the continuation value. Then we can define the per-period maximized value function as

$$\begin{aligned}\bar{V}^S_{j,g}(a_g, e_g) &= \max_{c,h} u^S_j(c,h) \\ \text{s.t.} \quad & c + p_j h = w_{j,e_g,g} a_g + T(w_{j,e_g,g} a_g).\end{aligned}$$

The housing demand associated with this maximization problem is given by

$$h = \frac{1-\alpha}{p_j}(w_{j,e_g,g} a_g + T(w_{j,e_g,g} a_g)) + \alpha \underline{h}.$$

3.5.2 Married households

Like singles, married households choose housing and consumption continuously. They also decide whether household members should participate in the labor market. Each period, at an arrival rate of δ, married households receive a shock to their marriage quality, and they take a new draw from ε^M. When this shock occurs the marriage survives if the condition in Equation (2) still holds. Otherwise, the marriage ends in divorce. We assume that there are no utility cost or side payments associated with a divorce. As a result, divorce follows the same threshold rule that is used for new marriages, and the probability of divorce is given by

$$\delta\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right).$$

The value of being married for a female is then given by

$$\begin{aligned}(r+\pi)V_{j,f}^M(a_m,e_m,a_f,e_f,\varepsilon^M) &= \max_{c,h,l_g\in\{0,1\}} u_j^M(c,h,l_f,l_m)+\varepsilon^M+ \\ &\delta\int_{-\infty}^{\bar{\varepsilon}(a_m,e_m,a_f,e_f)}(V_{j,f}^S(a_f,e_f)-V_{j,f}^M(a_m,e_m,a_f,e_f,\tilde{\varepsilon}))d\Phi\left(\frac{\tilde{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right),\end{aligned} \tag{4}$$

subject to

$$c+p_jh=I^M+T(I^M),\ \text{with}\ I^M=w_{j,e_m,m}a_ml_m+w_{j,e_f,f}a_fl_f.$$

As we did for singles, we can define the per-period maximized indirect value from consumption and housing as $\bar{V}_{j,f}^M(a_m,e_m,a_f,e_f)$ with the associated housing demand given by

$$h=\frac{1-\alpha}{p_j}(I^M+T(I^M))+\alpha\underline{h}.$$

3.5.3 Solving for the value functions

In Appendix B, we show that the value of being married is given by

$$\begin{aligned}V_j^M(a_m,e_m,a_f,e_f,\varepsilon^M) &= \frac{\bar{V}_j^M(a_m,e_m,a_f,e_f)}{r+\pi+\delta\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}+\frac{\varepsilon^M}{r+\pi} \\ &+\delta\left[\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\left(V_j^S(a_f,e_f)-\frac{\mu_{\varepsilon^M}}{r+\pi}\right)+\frac{\sigma_{\varepsilon^M}}{r+\pi}\phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\right].\end{aligned} \tag{5}$$

The value of being married has two parts. The first is the sum of the discounted indirect utility from consumption and housing and the marriage match quality, which households enjoy as long as the marriage lasts. The second line represents the value associated with the end of the marriage and captures the value of returning to singlehood.

Similarly, the value of being single can be written as

$$V_j^S(a_f, e_f) = \frac{1}{\Xi(a_f, e_f)} \bar{V}_j^S(a_f, e_f) +$$

$$\sum_{a_m, e_m} \frac{\xi_{j,f}(a_f, e_f a_m, e_f)}{\Xi(a_f, e_f)} \left[\left(1 - \Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\right) \frac{\bar{V}_M(a_f, e_f a_m, e_f) + \mu_{\varepsilon^M}}{r + \pi + \delta\Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)} + \frac{r + \pi + \delta}{r + \pi + \delta\Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)} \frac{\sigma_{\varepsilon^M}}{r + \pi} \phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right) \right], \quad (6)$$

where

$$\Xi(a_f, e_f) = (r + \pi)\left(1 + \sum_{a_m, e_m} \xi_{j,f}(a_f, e_f a_m, e_f) \frac{1 - \Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}{r + \pi + \delta\Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}\right).$$

Again, the value of being single has two components. The first term is the discounted sum of static utilities of being single. Discounting now depends on how quickly an individual matches and forms a marriage and how long such marriage lasts, which is determined by the marriage and divorce probabilities defined above. Upon marriage, households also obtain a utility gain, summarized by the second term in $V_j^S(a_f, e_f)$.

Given these value function, the threshold match quality $\bar{\varepsilon}_g(a_f, e_f, a_m, e_m)$ can be obtained as a solution to

$$V_j^M(a_f, e_f, a_m, e_m, \bar{\varepsilon}_g(a_f, e_f, a_m, e_m)) = V_j^S(a_g, e_g) \text{ for } g \in \{f, m\}.$$

which allows us to obtain the threshold match quality as the solution of the following non-linear equation

$$\begin{aligned} \bar{\varepsilon}_f(a_m, e_m, a_f, e_f) &= (r + \pi)\left(1 - \delta\Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\right) V_j^S(a_f, e_f) \\ &\quad - \frac{\delta}{r + \pi}\left(\sigma_{\varepsilon^M}\phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right) - \mu_{\varepsilon^M}\Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\right) \\ &\quad - \frac{\bar{V}_j^M(a_m, e_m, a_f, e_f)}{r + \pi + \delta\Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}. \end{aligned}$$

So far we have defined value functions and the associated thresholds for females. The problems for males can be defined similarly. Since a marriage occurs when both spouses profit from it, the match-specific marriage threshold is the maximum of thresholds for each gender

$$\bar{\varepsilon}(a_f, e_f, a_m, e_m) = \max\{\bar{\varepsilon}_m(a_f, e_f, a_m, e_m), \bar{\varepsilon}_f(a_f, e_f, a_m, e_m)\}. \quad (7)$$

3.6 Stationary distributions

In a stationary equilibrium, flows into and out of marriage has to balance each other in each city. Then if μ^S_{j,e_g,a_g} is the mass of single households of gender g is city j and $\mu^M_{j,e_f,a_f,e_f,a_f}$ that of married households, in a stationary equilibrium flows into marriage must be equal to flow into divorce for each type of marriages, i.e.,

$$\mu^S_{j,e_f,a_f}\xi_{j,f}(a_f, e_f a_m, e_f)\left(1-\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\right)=\left(\pi+\delta\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\right)\mu^M_{j,e_m,a_m,e_f,a_f}. \quad (8)$$

To measure the fraction of singles who have been married at least once in the past, i.e., the mass of divorced individuals, we can also define a similar condition that equates flows into divorce with outflows from the divorce states. This condition is given by

$$\sum_{e_m,a_m}\mu^M_{j,e_m,a_m,e_f,a_f}\delta\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)=\mu^D_{j,e_f,a_f}\left(\pi+\sum_{e_m,a_m}\xi_{j,f}(a_f, e_f a_m, e_f)\left(1-\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\right)\right),$$

where μ^D_{j,e_g,a_g} is the mass of gender g individuals of type (e, a) who are divorced.

3.7 Location and Education Decisions

Individuals start their lives as singles and decide in which city to live. Each city offers a discounted lifetime utility of being single there, $V_f(e_f, a_f)$ for women, as defined above. The location choices also depend on an idiosyncratic preferences for each city, denoted by ε^j. We assume that $\varepsilon^j \sim Gumbel(0, \sigma_\varepsilon)$ and, thus, its distribution is common across cities. Then, the location choice of a female with education e_f and ability a_f is determined by the following maximization problem.

$$V_f(e_f, a_f) = \max_{j\in\{0,1,\ldots,J\}} V^S_j(a_f, e_f) + \varepsilon^j. \quad (9)$$

We define $V_m(e_m, a_m)$ similarly. Let the associated location decision be given by $J_g(e, a)$.

Given $V_g(e_g, a_g)$, an individual decides whether to get a college education. Recall that all individual are born with some innate ability a, distributed with some function F_a. Also suppose the cost of education is distributed according to a logistic distribution with mean $\frac{\chi}{a^\eta}$ where a is ability. Hence, individuals with a higher ability face on average a lower utility costs of obtaining a college degree. Then, an agent will get an education if

$$V_g(1, a_g) - \left(\frac{\chi}{a_g^\eta} + \epsilon\right) \geq V_g(0, a_g), \quad (10)$$

where ϵ is drawn from a logistic distribution with mean zero and scale parameter σ_ϵ.

Let the associated education decision be denoted by $E_g(a)$.

3.8 Equilibrium

An equilibrium is a set of value function for married and single individuals, $V_j^M(a_f, e_f, a_m, e_m, \varepsilon^M)$ and $V_j^S(a_g, e_g)$, threshold match quality levels for each type of marriage, $\bar{\varepsilon}(a_f, e_f, a_m, e_m)$, the location and education decisions given by $J_g(e, a)$ and $E_g(a)$, the share (mass) of the population of each gender, skill and ability in each city, $\{\mu_{j,e,a,g}\}$, share of population of each gender, skill and ability in each city who is married and single, $\{\mu^M_{j,e_m,a_m,e_f,a_f}\}$ and $\{\mu^S_{j,e,a,g}\}$, wages in each city, $\{w_{j,e,g}\}$, and housing prices in each city $\{p_j\}$ such that:

1. The value function $V_j^M(a_f, e_f, a_m, e_m, \varepsilon^M)$ and $V_j^S(a_g, e_g)$ are given by equations (5) and (6).

2. The match quality threshold is given by equation (7).

3. Location decisions are given by (9).

4. Education decisions are determined by equation (10).

5. Wages are determined by firms' FOCs with

$$L_{j,e_f,f} = \sum_{a_f} a_f \left(\mu^S_{j,e_f,a_f,f} + \sum_{e_m,a_m} \mu^M_{j,e_m,a_m,e_f,a_f} l_f(e_m, a_m, e_f, a_f) \right)$$
$$L_{j,e_m,m} = \sum_{a_m} a_m \left(\mu^S_{j,e_m,a_m,m} + \sum_{e_f,a_f} \mu^M_{j,e_m,a_m,e_f,a_f} l_m(e_m, a_m, e_f, a_f) \right)$$

6. Aggregate share of skilled and unskilled individuals is consistent with education decisions.

7. $\{\mu_{j,e,a,g}\}$ are determined by individuals' location choice decisions, with $\sum_{j,e,a} \mu_{j,e,a,g} = \mu_g = 1$.

8. $\{\mu^M_{j,e_m,a_m,e_f,a_f}\}$ and $\{\mu^S_{j,e,a,g}\}$ are determined by individuals' marriage decisions, with

$$\mu_{j,e_f,a_f,f} = \sum_{e_m,a_m} \mu^M_{j,e_m,a_m,e_f,a_f} + \mu^S_{j,e_f,a_f,f}$$
$$\mu_{j,e_m,a_m,m} = \sum_{e_f,a_f} \mu^M_{j,e_m,a_m,e_f,a_f} + \mu^S_{j,e_m,a_m,m}$$

9. Housing market clear, i.e.,

$$\omega_j^0 p_j^{\omega_j^1} = \sum_{e,a,g} \mu_{j,e,a,g}^S h^S(p_j, I_{j,e,a,g}^S) + \sum_{e_f,a_f,e_m,a_m} \mu_{j,e_m,a_m,e_f,a_f}^M h^M(p_j, I_{j,e_m,a_m,e_f,a_f}^M),$$

where $I_{j,e,a,g}^S$ and I_{j,e_m,a_m,e_f,a_f}^M are the net incomes of single and married households and h^S and h^M are the corresponding housing demand functions.

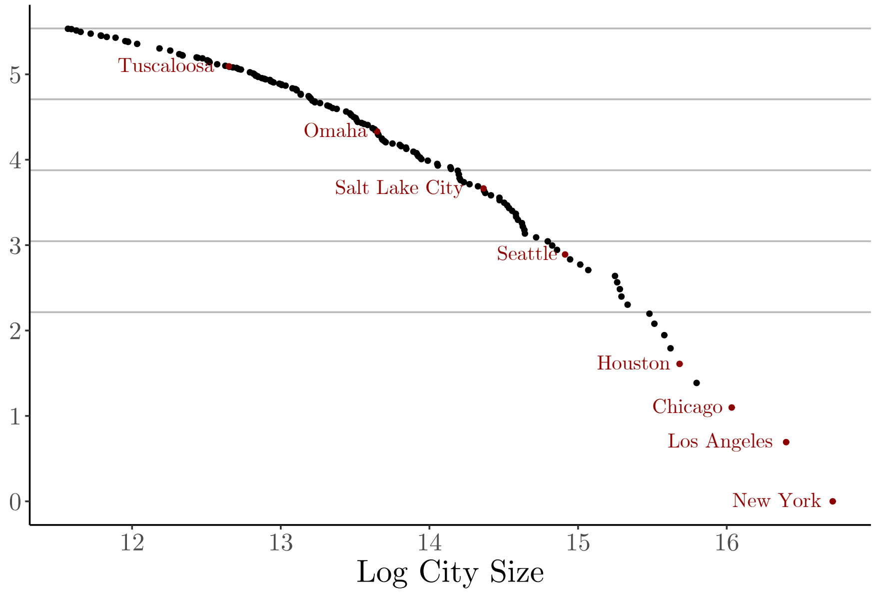
10. Marriage market is in equilibrium, i.e., equation (8) holds.

4 Benchmark Economy

We calibrate the model to the US economy in 2019. The first step in simulating the model economy is to define an empirical counterpart for the locations (cities) in the model. Our approach is to map US population living in MSAs into a small number of representative cities. In particular, we group MSAs into five cities based on their population in 2019. Since MSAs do not cover the entire US population, we define a residual rural location that groups counties that are not part of any metropolitan area. This implies that when an individual chooses a location (equation (9) in section 3.7), they consider six alternatives. We abstract from marriage and household decisions in the rural area, and calibrate the present expected value of living in the rural area for an individual of gender g and education e, $V_{g,e}^{rural}$, to match the fraction of individuals who live in rural areas in 2019.

Figure 9 plots the log population rank of MSAs in 2019 (vertical axis) and their corresponding log population (horizontal axis). The solid grey horizontal lines define the population thresholds that are used to group MSAs into five cities. Those lines, which coincide with breaks in the rank-population relationship, are defined as follows. Let thr_1 be the log population rank of the smallest MSA in 2019. This log rank, which in the data is given by $thr_1 = 5.537$ (population threshold 106,000), coincides with the lower rank threshold for the smallest city in the model (city 1). The bottom grey horizontal line, representing the lower rank threshold of the largest city (city 5), is instead given by $thr_5 = \frac{thr_1}{5} \times 2$ (population threshold 5,000,000). The rest of the lines are chosen to obtain four equally spaced groups, in terms of log population ranks, between thr_1 and thr_5. Therefore, the lower rank threshold for the second largest city (city 4) is given by $thr_4 = thr_5 + \frac{thr_1 - thr_5}{4}$, corresponding to population threshold 2,665,500. The lower rank threshold for the middle-sized city (city 3) is given by $thr_3 = thr_5 + \frac{thr_1 - thr_5}{4} \times 2$ (population threshold 1,455,000). The lower rank threshold for the second smallest city (city 2) is given by $thr_2 = thr_5 + \frac{thr_1 - thr_5}{4} \times 3$ (population threshold 506,000).

Figure 9: Log population rank of MSAs, 2019



Hence, city 1 in the model corresponds to all MSAs between thr_1 and thr_2, city 2 corresponds to all MSAs between thr_2 and thr_3, etc. Any model input or target we construct for a model city represents an average value across MSAs between the corresponding thresholds. Crucially, this procedure allows us to obtain 5 synthetic cities that can capture the empirical relationships between city-size and the other variables of interest.

Next, for each city in the model, we construct several data model inputs and data targets. The housing supply elasticities, ω_j^1, are borrowed from Baum-Snow and Han (2023). Saiz (2010) provides housing supply elasticity estimates using data from 1970 to 2000 for 237 metropolitan areas (MSAs). Baum-Snow and Han (2023) update Saiz (2010) estimation for the 2000-2010 period. For the benchmark economy, ω_j^1 values are set to their average values for the 2000-2010 period. Figure 10 (left panel) shows the relation between the city size and housing supply elasticities. In 2019, supply is less elastic in larger cities. The relation between city size and housing supply elasticity was similar in 2000. However, it was much more negative in 1980, when the elasticity in smaller cities tended to be substantially higher.

We measure amenities at the location-year level, B_j, following Diamond (2016). The amenities are constructed using MSA-level information on the extent and quality of the retail sector, the quality of the transportation network, the number of patents, crime rate, environmental quality, and schooling quality in 2019.[1] A single amenity index for each location and year is constructed using a principal component analysis.

[1]Data is based on the US Census, the County Business Patterns, the FBI Uniform Crime Reports, the EPA air quality monitors, the Census of Governments, the NBER Patent Database, the US Patent and Trademark Office, and Duranton and Turner (2011). Appendix A contains more details on the amenities used.

The right panel in Figure 10 shows how amenities differ across cities in the data. Not surprisingly, there is a very strong positive correlation between city size and estimated amenities.

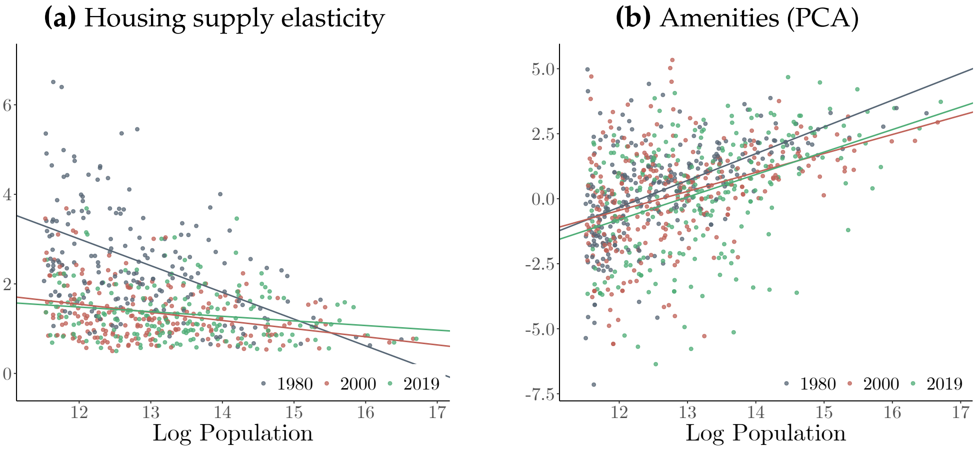


Figure 10: The figures plot MSA outcomes by year. Sources: Baum-Snow and Han (2023), and Saiz (2010) (left panel); Diamond (2016) and data sources referenced in footnote 1 (right panel).

A few parameters are set directly to their data counterparts or taken from the literature. The death probability, π, is chosen so that an average individual lives for 40 years, roughly the time spent in the labor market in the data. The rate of time preference, r, is set to 2%, which corresponds to an annual discount rate of 0.98. Since we cannot separately identify the average idiosyncratic value of marriage from the frequency of matches, we set the average efficiency of the matching function, ψ, to 0.151, a value estimated by Goussé, Jacquemet and Robin (2017). In a balanced marriage market, this value implies a median waiting time between two matches of 4.5 years. Finally, following Acemoglu and Autor (2011), we set the elasticity of substitution between skilled and unskilled labor, $1/(1-\rho)$, to 1.68.

Several parameters are chosen to exactly match certain targets *in equilibrium*. In particular, we choose X_j, the aggregate total factor productivity in city j, to match the average labor income in city j, and θ_j^N and κ_j^E to match the skill premium and the gender wage gap in each city in 2019. Data on labor income is from the 2019 American Community Survey (ACS), where we define two educational groups, $e \in \{N, E\}$, corresponding to individuals with and without a college degree. Finally, the scaling terms for housing supply, ω_j^0, are chosen so that in equilibrium we match average housing price in each city. In order to facilitate comparisons of housing prices across cities, we compute a city-level hedonic price index by adjusting for dwelling characteristics (details are given in Appendix A). All the targeted moments for a model city are calculated as the simple average of the corresponding moments for the MSAs that compose that particular city, as described above. The calibrated values for these parameters are reported in Table C1 in Appendix C.

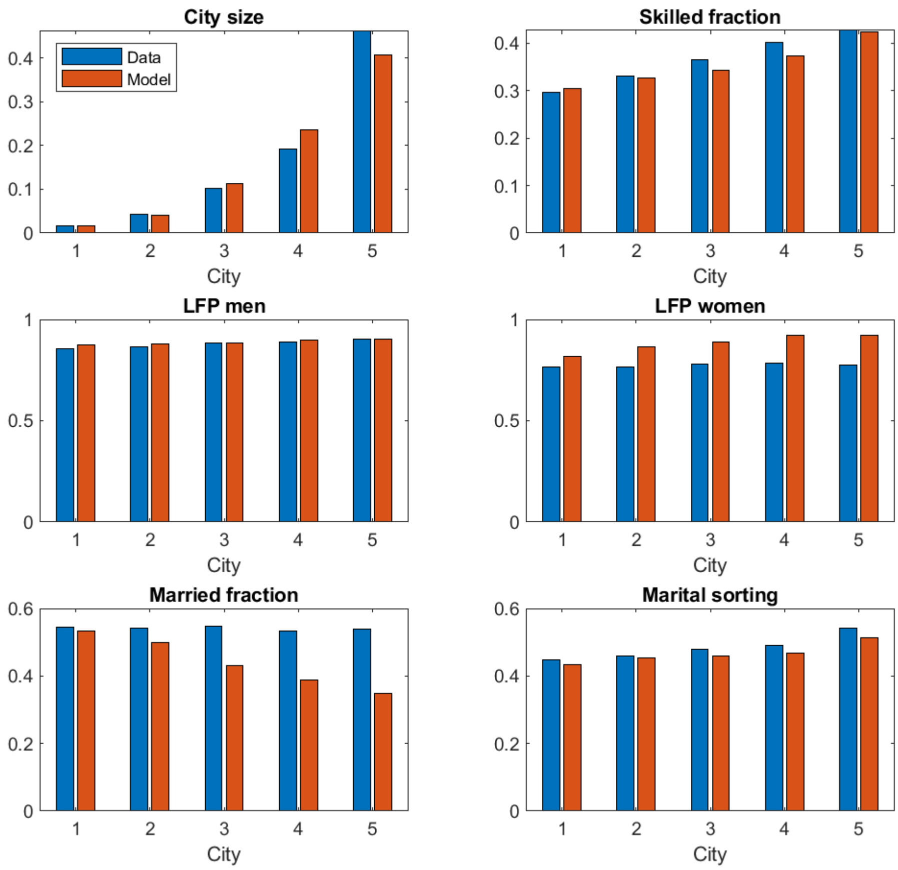


Figure 11: Benchmark Economy, Model vs. Data.

Moment	Model	Data
Population, skilled male, fraction	0.35	0.33
Population, skilled female, fraction	0.36	0.38
Housing, share total cost, 1st tercile of inc. distr.	0.46	0.42
Housing, share total cost, 2nd tercile of inc. distr.	0.35	0.37
Housing, share total cost, 3rd tercile of inc. distr.	0.30	0.33
Population, divorced fraction	0.08	0.10
Inequality, 90-10 perc. ratio	3.55	3.37

Table 1: Benchmark Economy, Model vs. Data.

We are then left with 21 parameters to determine, for $g \in \{f, m\}$, and $J = 5$: i) variance of the ability distribution, σ_a, ii) the average cost of education, χ, and heterogeneity in cost of education, σ_ϵ, iii) variance of preference shocks for locations, σ_ε, iv) cost of work, q_g, v) utility weight on consumption, α, vi) minimum housing consumption, $\underline{h}$, vii) marriage match quality, μ_{ε^M} and σ_{ε^M}, viii) divorce shock, δ, viii) additional probability of matching with your own education type, ϕ_j, ix) the utility weight on amenities, β, and x) value of living in a rural area, $V^{rural}_{g,e}$.

To estimate these parameters, we use a set of 37 targets:

1. Share of population in each city (5 targets)
2. Fraction of skilled population in each city (5 targets)
3. Labor force participation of men and women in each city (10 targets)
4. Fraction of population married in each city (5 targets)
5. Marital sorting, correlation of husbands and wives educations, in each city (5 targets)
6. Fraction of males and females with a college degree (2 targets)
7. Housing expenditure as a share of household total expenditure by terciles of income distribution (3 targets)
8. Fraction of population who are divorced (1 target)
9. Earnings inequality, 90-10 ratio (1 target).

Figure 11 and Table 1 shows model fit. Table 2 shows the estimated parameters.

5 Coming Apart

We next exploit the benchmark model to understand changes in US households. To this end, we conduct two counterfactual experiments. First, we change the parameters of the production function, X_j, θ^N_j and κ^e_j, so that average income, skill premium, and the gender wage gap in each city are equal to their 1980 values in the data. In both cases, we recalibrate the value of living in rural areas to match the shares of urban population observed in 1980. We keep all other parameters at their benchmark values. Table C2 in Appendix C shows the parameters of the production function for the 1980 economy. The counterfactual economy is characterized by lower skill premiums and a larger gender wage gaps. Moreover, in contrast to the benchmark economy, the gradient between city sizes and skill premium is lower. Second, we change the elasticity of the housing supply parameter, ω^1_j, to their 1980 values, as illustrated in Figure 12. In 1980, the housing supply elasticities were significantly higher in smaller cities. From 1980 to 2019, they have declined sharply in smaller cities and only so slightly in larger cities, such that by 2019 elasticities across cities are much similar than they were four decades earlier

Parameter	Value	Moments
Marriage, match quality, mean (ε^M)	-0.04	Fraction married by city
Marriage, match quality, std (σ_{ε^M})	0.27	
Marriage, match quality, shock prob (δ)	0.01	Divorced fraction
Match. fun. bias, city 1 (γ_j)	0.40	
Match. fun. bias, city 2 (γ_j)	0.38	
Match. fun. bias, city 3 (γ_j)	0.33	Assortative mating by city
Match. fun. bias, city 4 (γ_j)	0.28	
Match. fun. bias, city 5 (γ_j)	0.29	
Cost of Education (χ)	15.72	Share with college education
Cost of work, male (q_m)	0.61	LFP by gender
Cost of work, female (q_f)	0.43	
Ability distr., variance parameter (σ_a)	0.20	Aggregate income inequality
Utility weight on consumption (α)	0.74	Housing expenditure by income quintile
Minimum housing ($\underline{h}$)	0.57	
Amenities coefficient (β)	0.47	Correlation between amenities and city size
City pref., scale parameter (σ_ε)	8.71	Correlation between income and city size
Education shock, scale parameter (σ_ϵ)	8.23	Gender gap in education
Value rural, unskilled male	48.40	
Value rural, unskilled female	42.05	Share of rural population
Value rural, skilled male	52.25	
Value rural, skilled female	48.22	

Table 2: Estimated parameters.

As the wage structure changes between 1980 and 2019, the model generates a significant increase in the population share with a college degree, which increases from 20% in the 1980 counterfactual to 35% in 2019. The growth happens in all cities. But, as Figure 13 (left panel) shows, the largest city in the model, which already had the highest share of the skilled population in the 1980 counterfactual, experiences the largest increase in skilled population, consistent with great divergence observed in the data. With the decline in the gender wage gap and overall wage increase, the female labor force increases significantly between 1980 and 2019, as shown in Figure 13 (right panel). Again, consistent with the data, the increase is uniform across city size. While women enter the labor force, the labor force participation of less educated men declines (Figure 14).

Changes in wage structure have also affected the marital status of the population. The number of married people declines between 1980 and 2019, and more so in smaller cities, as shown in Figure 15 (left panel). As we observe in the data, there is also a significant increase in assortative mating, and more so in larger cities (Figure 15, right panel).

Finally, the higher geographic dispersion of productivity along with the higher skill premium, the higher concentration of skilled individuals, and the higher concentration of couples with two college graduates, generate a significant rise in the overall concentration of population in bigger cities which is accompanied by an increase in

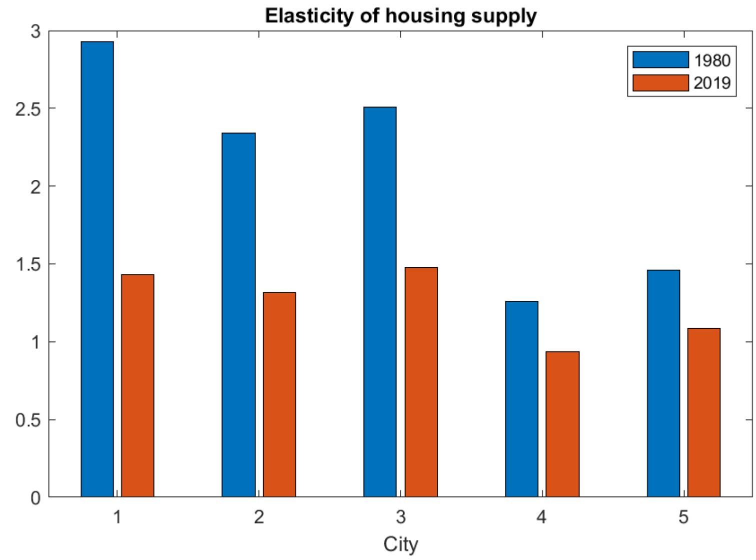


Figure 12: Elasticities of housing supply in 1980 and 2019.

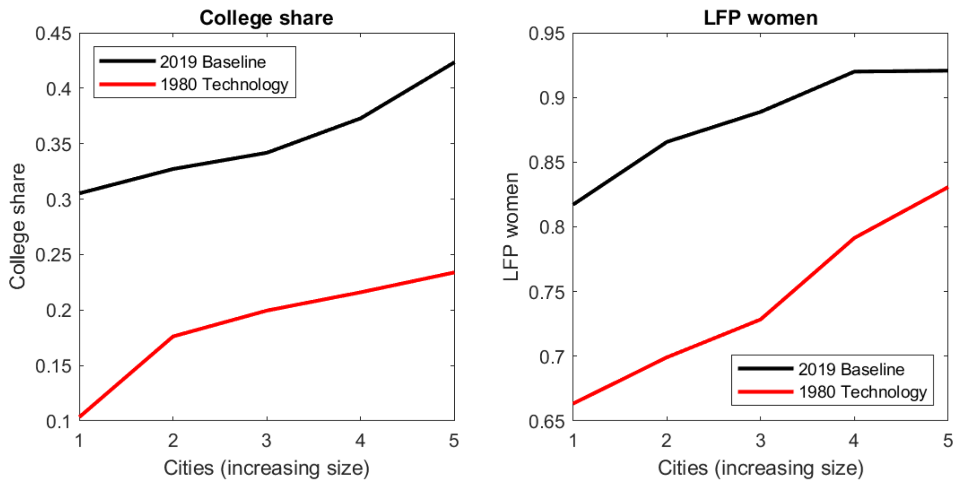


Figure 13: Skilled population share and married women labor force participation. Benchmark economy versus 1980 technology counterfactual.

housing prices (Figure 16). This causes more dispersion of house prices across space, in line with what has been observed in the US.

Next, we turn to the effects of changes in housing supply elasticities. Overall, as housing supply elasticities decline between 1980 and 2019, the house prices increase in all locations (Figure 17, right panel). Despite the uniformity of these changes the population shares increase more in bigger cities. These patterns are driven by significant changes in the composition of the population. Changes in housing prices, in fact, have implications for household structure. Higher housing prices cause an increase in female labor force participation, and the effect is larger in bigger cities, as shown in the left panel of Figure 18. Furthermore, for college graduates, higher house prices in larger cities make marrying another college graduate more attractive, making the

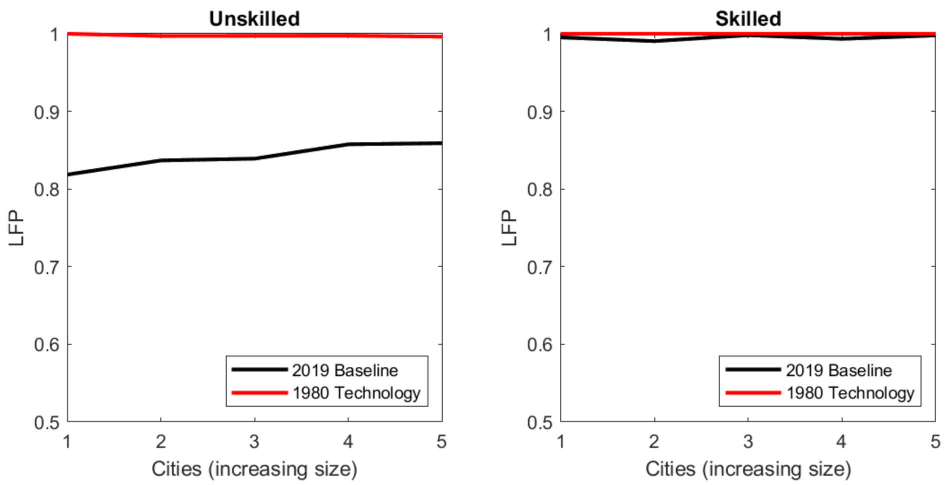


Figure 14: Labor force participation of men by skill level. Benchmark economy versus 1980 technology counterfactual.

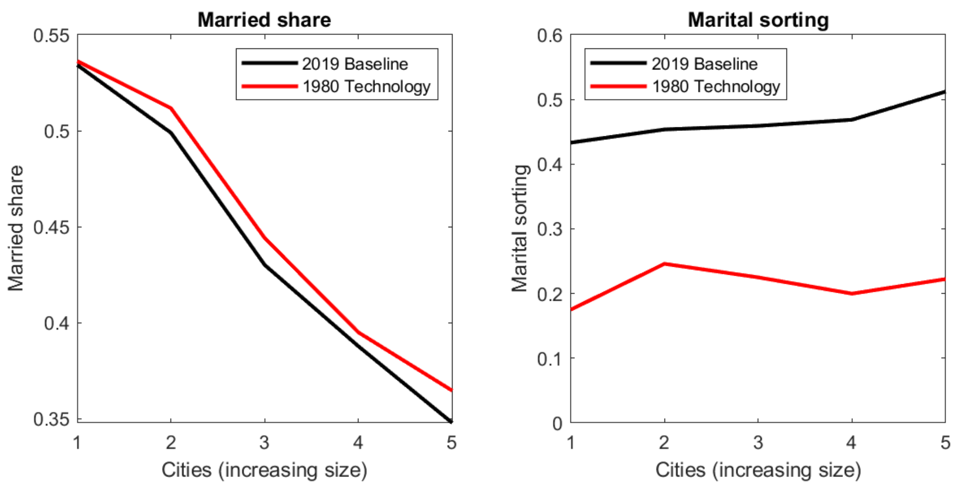


Figure 15: Marriage share and marital sorting. Benchmark economy versus 1980 technology counterfactual.

relation between city size and assortative mating more positive in the 2019 economy compared to the 1980 counterfactual (Figure 18, right panel).

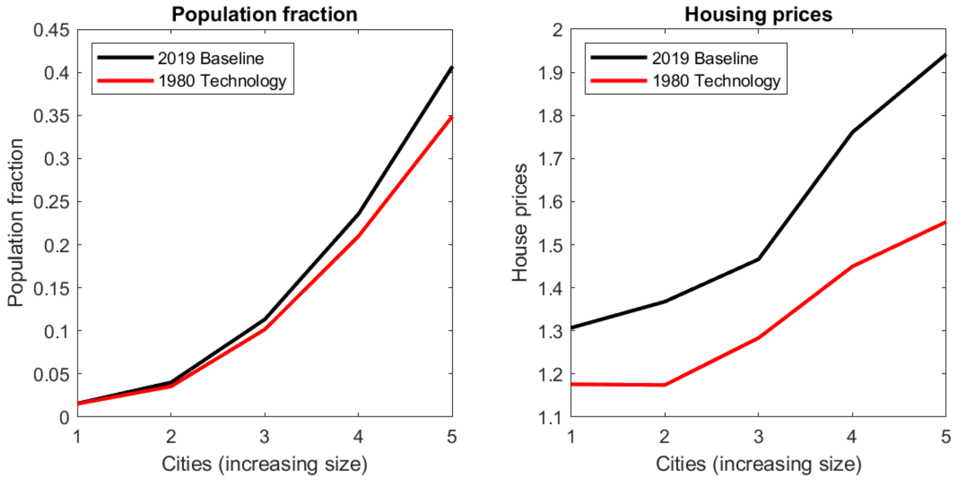


Figure 16: Population shares and housing prices. Benchmark economy versus 1980 technology counterfactual.

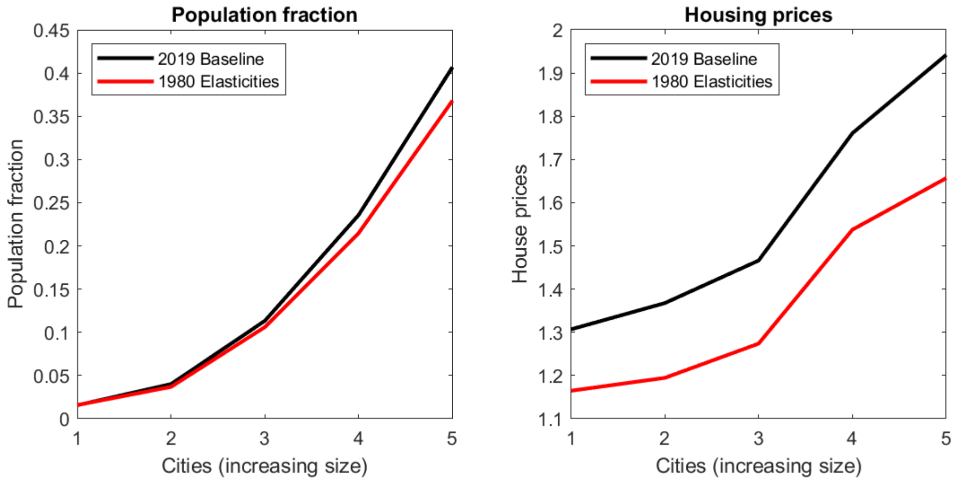


Figure 17: Population shares and housing prices. Benchmark economy versus 1980 elasticities of housing supply counterfactual.

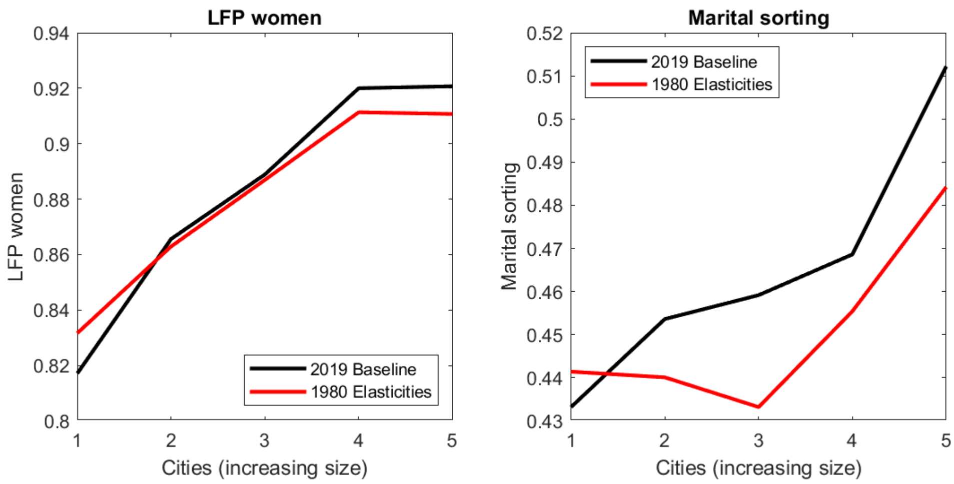


Figure 18: Married women labor force participation and marital sorting. Benchmark economy versus 1980 elasticities of housing supply counterfactual.

6 Conclusion

US households are much more segregated across space today than they were decades ago. Since the 1980s, there has been a significant concentration of college-educated individuals, particularly in larger cities where inequality has also risen. Smaller cities experienced a more substantial decline in marriage rates, especially among non-college graduates, leading to a doubling of single non-college graduates from 1980 to 2019. Assortative mating has notably increased in larger cities, contributing to higher inequality. Simultaneously, house price disparities across US cities have expanded. This societal shift is characterized by educational, marital, and geographical divisions.

To explore these trends, we employ a spatial heterogeneous agent model of family formation. The model reveals that changes in the wage structure between 1980 and 2019 concentrate more college graduates in larger cities, intensifying assortative mating. This also contributes to housing price disparities, with prices rising more significantly in larger cities. Additionally, the housing market changes, with a less elastic housing supply between 1980 and 2019, further fueling assortative mating in larger cities, as higher housing prices incentivize college graduates to marry individuals with similar educational backgrounds.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In . Vol. 4 of *Handbook of Labor Economics*, , ed. David Card and Orley Ashenfelter, 1043–1171. Elsevier. 3, 19

Alonzo, Davide. 2022. "Marrying Your Job: Matching and Mobility With Geographic Heterogeneity." Working Paper. 3

Baum-Snow, Nathaniel, and Lu Han. 2023. "The Microgeography of Housing Supply." *Journal of Political Economy*, Forthcoming. 18, 19, 30

Baum-Snow, Nathaniel, and Ronni Pavan. 2013. "Inequality and City Size." *The Review of Economics and Statistics*, 95(5): 1535–1548. 4

Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan. 2018. "Why Has Urban Inequality Increased?" *American Economic Journal: Applied Economics*, 10(4): 1–42. 4

Berry, Christopher R., and Edward L. Glaeser. 2005. "The Divergence of Human Capital Levels Across Cities." *Papers in Regional Science*, 84(3): 407–444. 3

Blasutto, Fabio. 2023. "Cohabitation VS. Marriage: Mating Strategies by Education in the Usa." *Journal of the European Economic Association*, jvad065. 2

Bénabou, Roland. 2002. "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica*, 70(2): 481–517. 11

Charles, Kerwin Kofi, Erik Hurst, and Mariel Schwartz. 2019. "The Transformation of Manufacturing and the Decline in US Employment." *NBER Macroeconomics Annual*, 33(1): 307–372. 3

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir. 2020. "Changes in Assortative Matching: Theory and Evidence for the US." NBER Working Paper 26932. 3

Compton, Janice, and Robert A Pollak. 2007. "Why Are Power Couples Increasingly Concentrated in Large Metropolitan Areas?" *Journal of Labor Economics*, 25(3): 475–512. 7

Costa, Dora L, and Matthew E Kahn. 2000. "Power Couples: Changes in the Locational Choice of the College Educated, 1940–1990." *The Quarterly Journal of Economics*, 115(4): 1287–1315. 7

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of Us Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review*, 106(3): 479–524. 2, 18, 19, 30

Duranton, Gilles, and Matthew A Turner. 2011. "The Fundamental Law of Road Congestion: Evidence From Us Cities." *American Economic Review*, 101(6): 2616–2652. 18, 30

Eckert, Fabian, Sharat Ganapati, and Conor Walsh. 2022. "Urban-Biased Growth: A Macroeconomic Analysis." NBER Working Paper 30515. 2

Fan, Jingting, and Ben Zou. 2022. "The Dual Local Markets: Family, Jobs, and the Spatial Distribution of Skills." Working Paper. 3

Giannone, Elisa. 2022. "Skilled-Biased Technical Change and Regional Convergence." Working Paper. 2

Goussé, Marion, Nicolas Jacquemet, and Jean-Marc Robin. 2017. "Marriage, Labor Supply, and Home Production." *Econometrica*, 85(6): 1873–1919. 2, 11, 19

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 2014. "Marry Your Like: Assortative Mating and Income Inequality." *American Economic Review*, 104(5): 348–53. 3

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 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. 2

Guner, Nezih, Martin Lopez-Daneri, and Gustavo Ventura. 2023. "The Looming Fiscal Reckoning: Tax Distortions, Top Earners, and Revenues." *Review of Economic Dynamics*, Forthcoming. Special Issue in Memory of Tom Cooley. 11

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006." *Review of Economic Dynamics*, 13(1): 15–51. Special issue: Cross-Sectional Facts for Macroeconomists. 3

Heathcote, Jonathan, Fabrizio Perri, Giovanni L. Violante, and Lichen Zhang. 2023. "More Unequal We Stand? Inequality Dynamics in the United States, 1967–2021." *Review of Economic Dynamics*, 50: 235–266. 3

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante. 2014. "Consumption and Labor Supply With Partial Insurance: An Analytical Framework." *American Economic Review*, 104(7): 2075–2126. 11

Krueger, Alan. 2017. "Where Have All the Workers Gone? An Inquiry into the Decline of the U.S. Labor Force Participation Rate." *Brookings Papers on Economic Activity*, 48(2): 1–87. 3

Moreno-Maldonado, Ana. 2023. "Mums and the City: Household Labour Supply and Location Choice." Working Paper. 2

Moreno-Maldonado, Ana, and Clara Santamaria. 2022. "Delayed Childbearing and Urban Revival." Working Paper. 3

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." *The Quarterly Journal of Economics*, 125(3): 1253–1296. 18, 19, 30

Van Nieuwerburgh, Stijn, and Pierre-Olivier Weill. 2010. "Why Has House Price Dispersion Gone Up?" *The Review of Economic Studies*, 77(4): 1567–1606. 5

Appendix

This Appendix is organized as follows. Section A provides details about the data. Section B presents further details about the model. Section C contains additional tables referenced in the main text.

A Data

[Back]

Data on marital status, employment, wages, housing cost, and education, comes from the 5 percent samples of the US Census 1980, 2000, and the ACS 5-year 2015-2019 sample. We focus on individuals aged 25 to 54. We group them by educational attainment "college degree" versus "no college degree".

The local inverse housing supply elasticities for the years 1960 and 1980 are sourced from Saiz (2010), while those for the more recent decades (2000 and 2019) come from Baum-Snow and Han (2023). Saiz (2010) reports elasticity estimates using data from 1970 to 2000 for 237 metropolitan areas (MSAs), whereas Baum-Snow and Han (2023) replicate Saiz' estimation for the 2000-2010 period. Elasticity estimates within MSAs are fixed to their 1970-2000 values in 1960 and 1980, and to their 2000-2010 values in 2000 and 2019.

We measure amenities at the location-year level following Diamond (2016). Information on amenities related to retail, transportation, patents, crime, environmental quality, and schooling quality is collected from multiple data sources: The US Census, the County Business Patterns, the FBI Uniform Crime Reports, the EPA air quality monitors, the Census of Governments, the NBER Patent Database, the US Patent and Trademark Office, and Duranton and Turner (2011). Retail amenities reflect the variety of shops and entertainment options in cities, and are measured by the per capita number of clothing stores, eating and drinking places, and cinemas. Transportation amenities assess the quality of public transit and road infrastructure, including data on buses per person, an overall public transit rating, and average daily traffic on major roads. Patents per capita are also measured in the data. Crime data cover both violent and property crimes per person. Environmental amenities include per capita government spending on parks and recreation, as well as the EPA's air quality index. School quality measures include government spending per student in K–12 education and the number of students enrolled in private schools relative to those in public schools.

A single amenity index for each location and year is constructed using a principal component analysis, as in Diamond (2016). Relative to Diamond, our index is also computed for the year 2019. Due to the unavailability of transportation data for that year, we rely on data from 2010 to measure transportation amenities.

Hedonic Price Index In order to facilitate comparisons of housing prices across cities, we calculate a hedonic price index for each MSA in the sample. Our approach involves regressing the log gross annual rent paid by each renter i on a set of location-specific fixed effects and housing characteristics for each year t. The housing characteristics include the number of bedrooms, rooms, household members per room, and the total number of units in the structure in which the individual resides.

Our regression equation for individual i and given year t is expressed as follows:

$$\log(p_i) = \alpha_{\text{MSA}(i)} + \alpha_1 \text{Rooms}_i + \alpha_2 \text{Units}_i + \alpha_3 \text{Bedrooms}_i + \alpha_4 \left(\frac{\text{members}_i}{\text{rooms}_i} \right) + \varepsilon_i.$$

The reference groups for each cathegorical variable is given by the median value at the national level for each year.

The hedonic price index for each MSA is given by the fixed-effect $\alpha_{\text{MSA}(i)}$ obtained from this regression. The final prices used in the model, p_j, are expressed in thousand of euros, and are obtained by taking the exponential of $\alpha_{\text{MSA}(i)}$, dividing by 1,000, and averaging within city groups. Given the way the regression is estimated, one unit of housing refers to a dwelling having the median number of rooms, units in the structure, bedrooms, and members per room observed in the U.S. for that specific year.

B Solving for the Value Functions

[Back]

In this appendix, we derive equations (5) and (6). In the derivations, to enhance legibility, we omit all the indexes regarding abilities and education whenever possible (e.g. $V_{j,f}^M(a_m, e_m, a_f, e_f, \varepsilon^M)$ becomes $V_{j,f}^M(\varepsilon^M)$). First, we notice that (4) can be rewritten as

$$V_{j,f}^M(\varepsilon^M) - \frac{\varepsilon^M}{r+\pi} = \frac{\bar{V}_j^M}{r+\pi} + \frac{\delta}{r+\pi} \int_{-\infty}^{\bar{\varepsilon}} (V_{j,f}^S - V_{j,f}^M(\tilde{\varepsilon})) d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right), \tag{11}$$

which implies that

$$V_{j,f}^M(\varepsilon^M) - \frac{\varepsilon^M}{r+\pi} = V_{j,f}^M(\hat{\varepsilon}) - \frac{\hat{\varepsilon}}{r+\pi}$$

for any $\hat{\epsilon}$. Using this result, we can rewrite the continuation value in (11) as

$$\begin{aligned}
\int_{-\infty}^{\bar{\varepsilon}} (V_{j,f}^S - V_{j,f}^M(\tilde{\varepsilon})) d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) &= \int_{-\infty}^{\bar{\varepsilon}} \left(V_{j,f}^S - V_{j,f}^M(\tilde{\varepsilon}) + \frac{\tilde{\varepsilon}}{r+\pi} - \frac{\tilde{\varepsilon}}{r+\pi} \right) d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \\
&= \int_{-\infty}^{\bar{\varepsilon}} \left(V_{j,f}^S - V_{j,f}^M(\hat{\varepsilon}) + \frac{\hat{\varepsilon}}{r+\pi} - \frac{\tilde{\varepsilon}}{r+\pi} \right) d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \\
&= \Phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \left(V_{j,f}^S - V_{j,f}^M(\hat{\varepsilon}) + \frac{\hat{\varepsilon}}{r+\pi} \right) - \\
&\quad \frac{1}{r+\pi} \int_{-\infty}^{\bar{\varepsilon}} \tilde{\varepsilon} d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right)
\end{aligned}$$

For $\hat{\varepsilon} = \varepsilon^M$ and since

$$\int_{-\infty}^{\bar{\varepsilon}} \tilde{\varepsilon} d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) = \Phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \mu_{\varepsilon^M} - \phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \sigma_{\varepsilon^M}$$

we have

$$\int_{-\infty}^{\bar{\varepsilon}} (V_{j,f}^S - V_{j,f}^M(\tilde{\varepsilon})) d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) = \Phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \left(V_{j,f}^S - V_{j,f}^M(\varepsilon^M) + \frac{\varepsilon^M - \mu_{\varepsilon^M}}{r + \pi} \right) + \frac{\sigma_{\varepsilon^M}}{r + \pi} \phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right)$$

Replacing the latter in the value function for married female and rearranging gives

$$\begin{aligned} V_j^M(\varepsilon^M) &= \frac{\bar{V}_j^M}{r + \pi + \delta \Phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right)} + \frac{\varepsilon^M}{r + \pi} \\ &\quad + \delta \left[\Phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) (V_j^S - \frac{\mu_{\varepsilon^M}}{r + \pi}) + \frac{\sigma_{\varepsilon^M}}{r + \pi} \phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \right] \end{aligned}$$

which is equation (5) in the main text. To derive equation (6), first notice that using the latter equation we can rewrite the integral term in the continuation value of equation (3) as

$$\int_{\bar{\varepsilon}}^{\infty} \left[V_j^M(\tilde{\varepsilon}) - V_j^S \right] d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) =$$
$$\frac{1 - \Phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right)}{r + \pi + \delta \Phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right)} \left[\bar{V}_j^M - (r + \pi) V_j^S - \frac{\delta}{r + \pi} \left(\mu_{\varepsilon^M} \Phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) - \sigma_{\varepsilon^M} \phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \right) \right] + + \frac{1}{r + \pi} \int_{\bar{\varepsilon}}^{\infty} \tilde{\varepsilon} d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right)$$

Since

$$\int_{\bar{\varepsilon}}^{\infty} \tilde{\varepsilon} d\Phi \left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) = \left(1 - \Phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \right) \mu_{\varepsilon^M} + \phi \left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}} \right) \sigma_{\varepsilon^M}$$

we have

$$\int_{\bar{\varepsilon}}^{\infty} \left[V_j^M(\tilde{\varepsilon}) - V_j^S\right] d\Phi\left(\frac{\tilde{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right) =$$

$$\frac{1 - \Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}{r + \pi + \delta\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}\left[\bar{V}_j^M + \mu_{\varepsilon^M} - (r+\pi)V_j^S\right] + \frac{r+\pi+\delta}{r + \pi + \delta\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}\frac{\sigma_{\varepsilon^M}}{r+\pi}\phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)$$

Finally, using the latter in the value function for singles and rearranging gives

$$V_j^S = \frac{1}{\Xi}\bar{V}_j^S + \sum_{a_m, e_m}\frac{\xi_{j,f}}{\Xi}\left[\left(1 - \Phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\right)\frac{\bar{V}_M + \mu_{\varepsilon^M}}{r + \pi + \delta\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)} + \frac{r+\pi+\delta}{r + \pi + \delta\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}\frac{\sigma_{\varepsilon^M}}{r+\pi}\phi\left(\frac{\bar{\varepsilon} - \mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)\right]$$

where

$$\Xi = (r+\pi)\left(1 + \sum_{a_m, e_m}\xi_{j,f}\frac{1 - \Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}{r + \pi + \delta\Phi\left(\frac{\bar{\varepsilon}-\mu_{\varepsilon^M}}{\sigma_{\varepsilon^M}}\right)}\right)$$

which is equation (6).

C Additional Tables

[Back]

Parameter	Value	Moments
Prod. fun., TFP, city 1 (X_j)	12.67	
Prod. fun., TFP, city 2 (X_j)	13.69	
Prod. fun., TFP, city 3 (X_j)	14.15	Average income by city
Prod. fun., TFP, city 4 (X_j)	15.33	
Prod. fun., TFP, city 5 (X_j)	16.96	
Prod. fun., skill share, city 1 ($1-\theta_j^N$)	0.53	
Prod. fun., skill share, city 2 ($1-\theta_j^N$)	0.54	
Prod. fun., skill share, city 3 ($1-\theta_j^N$)	0.56	Skill premium by city
Prod. fun., skill share, city 4 ($1-\theta_j^N$)	0.58	
Prod. fun., skill share, city 5 ($1-\theta_j^N$)	0.64	
Prod. fun., skilled gender gap, city 1 (κ_j^E)	0.79	
Prod. fun., skilled gender gap, city 2 (κ_j^E)	0.72	
Prod. fun., skilled gender gap, city 3 (κ_j^E)	0.70	Gender wage gap by city, skilled
Prod. fun., skilled gender gap, city 4 (κ_j^E)	0.73	
Prod. fun., skilled gender gap, city 5 (κ_j^E)	0.71	
Prod. fun., unskilled gender gap, city 1 (κ_j^N)	0.80	
Prod. fun., unskilled gender gap, city 2 (κ_j^N)	0.76	
Prod. fun., unskilled gender gap, city 3 (κ_j^N)	0.77	Gender wage gap by city, unskilled
Prod. fun., unskilled gender gap, city 4 (κ_j^N)	0.78	
Prod. fun., unskilled gender gap, city 5 (κ_j^N)	0.78	
Housing supply intercept, city 1	0.14	
Housing supply intercept, city 2	0.35	
Housing supply intercept, city 3	0.85	Housing prices by city
Housing supply intercept, city 4	1.74	
Housing supply intercept, city 5	2.52	

Table C1: Production parameters. [Back]

TFP					
Year	City 1	City 2	City 3	City 4	City 5
2019 (baseline)	12.67	13.69	14.15	15.33	16.96
1980 (counterfactual)	10.73	11.12	11.25	11.76	12.56
Share skilled in production $1-\theta_j^N$					
Year	City 1	City 2	City 3	City 4	City 5
2019 (baseline)	0.53	0.54	0.56	0.58	0.64
1980 (counterfactual)	0.33	0.36	0.39	0.39	0.42
Gender gap skilled					
Year	City 1	City 2	City 3	City 4	City 5
2019 (baseline)	0.79	0.72	0.70	0.73	0.71
1980 (counterfactual)	0.48	0.59	0.56	0.58	0.59
Gender gap unskilled					
Year	City 1	City 2	City 3	City 4	City 5
2019 (baseline)	0.80	0.76	0.77	0.78	0.78
1980 (counterfactual)	0.58	0.57	0.59	0.59	0.61

Table C2: Production parameters: baseline and 1980 technology counterfactual. [Back]