Measuring and comparing assortativeness

Lecture material from Chiappori, Costa-Dias and Meghir, “Changes in Assortative Matching: Theory and Evidence for the US”

Mónica Costa-Dias

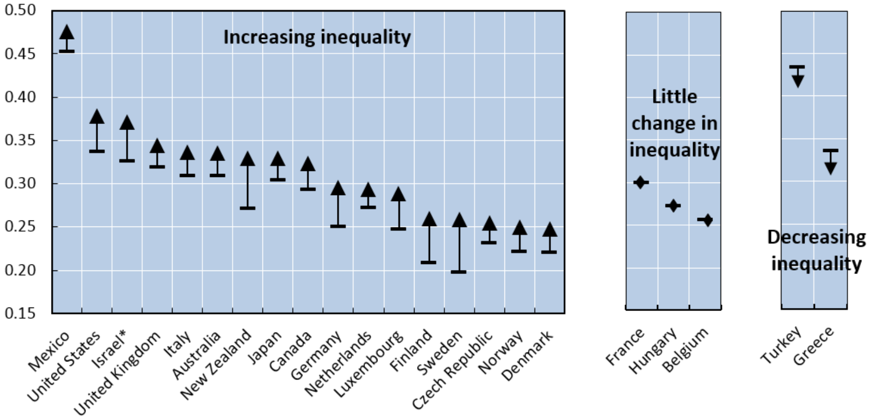
University of Bristol and Institute for Fiscal Studies

Econometric Society DSE Summer School

Bonn, August 2021

Generalised rise in inequality over the recent past

Gini coefficients for family disposable income, mid 80s to late 00s



Marital sorting likely important

- Refers to the extent to which **like marries like**
- Likely driven by benefits of marrying someone similar or different
- If traits like **education**, **skills** or **market productivity** matter for sorting, marriage is likely to magnify income inequality
- The higher concentration of resources at the family level accentuates inequality in resources available to children, carrying long-term intergenerational effects

Key questions

Here we focus on **marital sorting by education**

* Did assortative matching by education increase in the US?
* How would marital patterns look like today if assortativeness had not changed?
* To what extent did changes in marital sorting affect inequality in family earned income?

To address this question we need to establish

* How to define assortativeness
* How to define changes in assortativeness
* How to measure these concepts
* How to reflect changes in assortativeness into marital patterns

Measuring assortativeness and changes therein

- Detecting positive or negative assortativeness is not difficult
- But quantifying changes in sorting can be challenging
- *Key difficulty is to separate mechanical effects of changes in marginal distributions from deeper structural changes*
    - * Sorting depends on structure of gains from marrying different types
    - * Changes in sorting depend on how this structure changes

  → Establishing changes in sorting without reference to a model is difficult

Did assortativeness increase in the US?

Past studies reached contradictory conclusions

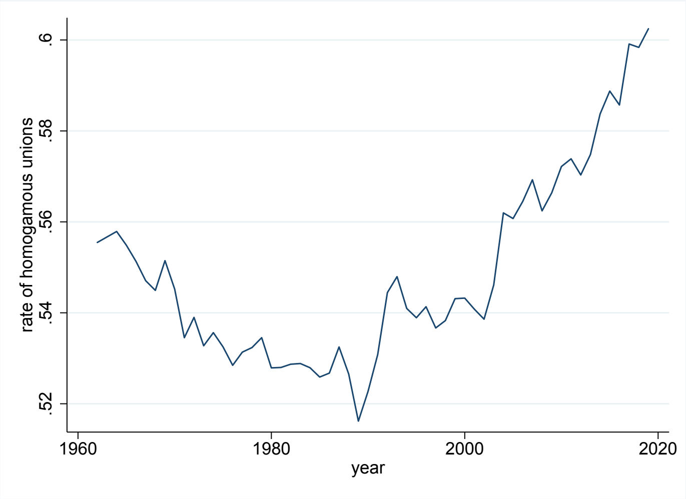
* Schwartz & Mare 2005, Greenwood et al 2014, Chiappori et al. 2017: assortativeness by education increased in the US over the past 50 years

* Eika et al. 19: increased at the bottom & declined at the top of the education distribution

* Shen 19: decreased up to 2000 and then increased

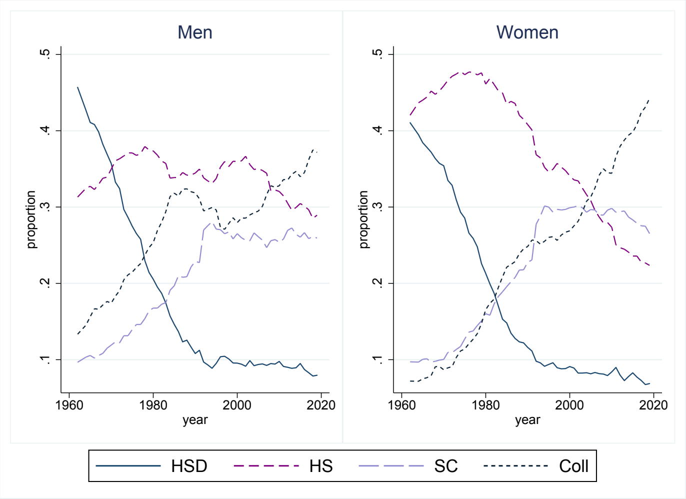
Trend in proportion of homogamous couples suggests increase in assortativeness since 80s

Proportion of couples with equally educated spouses



But may result from changes in education

Distribution of education by gender, 35-44 years old



What we do

1. Define assortativeness and changes in assortativeness
   * Propose minimal set of properties a measure of assortativeness should satisfy to detect changes in assortativeness
2. Propose new measure of assortativeness that relates to the economic structure of the matching problem
3. Develop its empirical counterpart in the Choo-Siow (2006) framework
4. Propose new, distribution-free concept
5. Discuss other measures of assortativeness
6. Show how assortativeness by education changed in the US
7. Use structural model to quantify role of changes in assortativeness on marital patterns

Meaning of assortativeness and changes therein
Setup

- Finite number of observed types
- Simple 2×2 sorting matrices
  - No loss of generality: measures of sorting are local
  - Overall picture pieced together

    Assortativeness can be +ve (or increasing) at one end and -ve (or decreasing) at another

  - Extend to higher order matching tables later
- 2 education groups, say college (C) and high school (HS) graduates
- Conditional on marriage: discuss singles later

Matching table

Table (m, n, r)

Men \ Women	C	HS
C	r	$m - r$
HS	$n - r$	$1 - n - m + r$

- m (n) proportion of higher educated men (women)
- r proportion of couples where spouses are both college educated

Definition of Positive Assortative Matching

- Benchmark: random matching
- PAM: *more people on diagonal* than under random matching
- In practice, the sorting matrix under random matching is

Table (m, n, mn)

Men \ Women	C	HS
C	mn	$m(1-n)$
HS	$n(1-m)$	$(1-m)(1-n)$

- This implies

$$
\begin{aligned}
\text{PAM} \quad &\Leftrightarrow \quad r \geq mn \\
&\Leftrightarrow \quad 1-n-m+r \geq (1-n)(1-m)
\end{aligned}
$$

Comparing assortative matching

* Comparing Table (m, n, r) & Table (m', n', r') is much harder
* Consider more PAM – all similar for NAM
* We start with two simple cases

Comparing assortative matching: special cases

1. **Same marginals:** $m = m'$ and $n = n'$

   (m, n, r) more AM than (m, n, r') iff more people in diagonal

   Formally: $r \geq r'$ or $1 - n - m + r \geq 1 - n - m + r'$

2. **Perfect assortative matching:** $m = n = r$

   AM is maximal for Table (m, m, m)

Men \ Women	C	HS
C	m	0
HS	0	$1 - m$

Formal requirements that an index of assortativeness should satisfy

- Usual practice: use index of assortativeness $I(m, n, r)$ to quantify degree of assortativeness and changes thereof

$$(m, n, r) \text{ more assortative than } (m', n', r')$$

$$\text{iff } I(m, n, r) \geq I(m', n', r')$$

- $I(m, n, r)$ must satisfy the conditions

1. *Monotonicity:* $I(m, n, r)$ increasing in r
2. *Perfectly Assortative Matching:* $I(m, m, m) \geq I(m', n', r')$ for all (m, m', n', r')

- Note that this characterisation *is not complete*
- So alternative indices of assortativeness satisfying these properties may reach different empirical results

Matching under Transferable Utility

- Relate sorting to underlying forces: *gains from matching together instead of matching another or remain single*

- Population of men i with characteristics X_i and women j with characteristics Y_j
  - * X and Y matter for the value of marriage, desirable

- TU: economic value of match between man i and woman j is

$$S(X_i, Y_j) \; = \; U(X_i, Y_j) + V(X_i, Y_j)$$

  - * Normalise value of singlehood: $U(X_i, 0) = V(0, Y_j) = 0$

$\rightarrow$ $S(X_i, Y_j)$ is the surplus of the match (increasing in X & Y)

Supermodularity and Assortativeness

- 2-dim surplus function $S(X, Y)$ is supermodular if for all $X \leq X'$, $Y \leq Y'$

$$S(X, Y) + S(X', Y') - S(X, Y') - S(X', Y) \geq 0$$

- Implies complementarity in spouses characteristics: increase in surplus from replacing Y with Y' increases with X:

$$S(X', Y') - S(X', Y) \geq S(X, Y') - S(X, Y)$$

- Supermodularity creates the incentive to sort positively in the marriage market

Structural Index of Assortativeness under TU

Table: Surplus matrix for matching between C and HS

Men \ Women	C	HS
C	$S(C,C)$	$S(C,HS)$
HS	$S(HS,C)$	$S(HS,HS)$

Structural index of assortativeness is supermodular core of this matrix

$$SM \;=\; S(C,C) + S(HS,HS) - S(C,HS) - S(HS,C)$$

* $SM > 0$ implies a tendency to PAM
* The more positive the core, the stronger the incentive
* Value of singlehood does not matter for this measure, what matters is how the gains from marrying up vary with education
* However, value of singlehood does determine probability of marriage

Empirical specification under Choo-Siow approach

- SM is impractical since values of marriage are not observed
- Choo-Siow (2006) offers structural interpretation of assortativeness, explicitly relating sorting patterns to gains from marriage in equilibrium

$\rightarrow$ Allows us to map changes in sorting to changes in surplus

$\rightarrow$ Compare surplus of match in different equilibria to say which is more assortative

- It will also allow us to consider unobserved heterogeneity in preferences

Choo-Siow framework

- Frictionless marriage market under TU
- Endogenous sharing rule clears the market
  - * Mr i will select spouse that maximises his utility
  - → u_i is price any woman has to pay to marry him
  - → Sharing rule is $S_{ij} = u_i + v_j$

Choo Siow Assumptions

- Large number of men and women participate in market
- Small number (N) of observed types (e.g. education)
- Surplus of match between Mr $i \in I$ and Mrs $j \in J$

$$S_{ij} = Z^{IJ} + \gamma_{ij}$$

Z^{IJ}: economic value of match (I, J)
γ_{ij}: unobserved preferences for each other

- Random term γ_{ij} is separable: $\gamma_{ij} = \varepsilon_i^J + \nu_j^I$, with ε and ν are independent from each other
- ε and ν: type 1 extreme value (EV1)

Choo-Siow (2006) result

- **Result:** There exist $\left(U^{IJ}, V^{IJ}\right)$ with $Z^{IJ} = U^{IJ} + V^{IJ}$ for all (I, J) such that, for matched pairs in equilibrium

$$u_i = U^{IJ} + \varepsilon_i^J \quad \text{and} \quad v_j = V^{IJ} + \nu_j^I$$

- Implies that i chooses a spouse in J iff

$$U^{IJ} + \varepsilon_i^J \geq U^{IK} + \varepsilon_i^K \text{ for all } K = 0, \ldots, N$$

- Under the EV assumption, $\left(U^{IJ}, V^{IJ}\right)$ can be retrieved

$$P^{IJ} = \frac{\exp\{U^{IJ}\}}{\sum_K \exp\{U^{IK}\}} \quad \text{and} \quad Q^{IJ} = \frac{\exp\{V^{IJ}\}}{\sum_K \exp\{V^{IK}\}}$$

The ‘Separable Extreme Value’ Index

- $\left(P^{IJ}, Q^{IJ}\right)$ are observed. In our example

$$r \;=\; n \times P^{CC} \;=\; m \times Q^{CC}$$

- Recall structural index of assortativeness is the supermodular core of the surplus matrix. Here only its systematic part matters:

$$SM \;=\; Z^{CC} + Z^{HS,HS} - Z^{C,HS} - Z^{HS,C}$$

- Under separability and EV, we call it I_{SEV}

$$I_{SEV}(m,n,r) \;=\; \ln\left(\frac{r\,(1+r-m-n)}{(n-r)\,(m-r)}\right)$$

- Table (m, n, r) is more assortative than Table (m', n', r') in the SEV sense iff

$$I_{SEV}(m, n, r) \geq I_{SEV}(m', n', r')$$

- Assortativeness directly related to supermodularity of the deterministic component

- Separability implies no stochastic super- (or sub-) modularity

Properties of I_{SEV}

- (Monitonicity) $I_{SEV}(m, n, r)$ is increasing in r
- (Perfectly assortative matching) $I_{SEV}(m, m, m) = +\infty$
- $I_{SEV}(m, n, mn) = 0$ and so PAM $\Leftrightarrow$ $I_{SEV}(m, n, r) \geq 0$
- I_{SEV} does not depend on probability of singlehood
- But recovering the strutural value of marriage does require the whole marital information, including number of singles

Relaxing the EV assumption

- EV assumption not needed for one-to-one relationship between the structure of choice and sorting patterns (Galichon-Salanie 2017)
- Under separability, any distribution can be used to derive an index of assortativeness that is the supermodular core of surplus table
- But comparing matching tables requires the same underlying of distribution of stochastic preferences
- And different distributional assumptions may lead different rankings of assortativeness

The generalised separable approach

- SEV conclusions regarding variations in assortativeness depend on EV assumption, which is not testable

- **Result** (m, n, r) more assortative than (m', n', r') for all possible distributions *iff*

$$\frac{r}{m} \geq \frac{r'}{m'} \quad \text{and} \quad \frac{r}{n} \geq \frac{r'}{n'} \quad \text{and}$$
$$\frac{1-n-m+r}{1-m} \geq \frac{1-n'-m'+r'}{1-m'} \quad \text{and}$$
$$\frac{1-n-m+r}{1-n} \geq \frac{1-n'-m'+r'}{1-n'}$$

- Unambiguous increase in assortativeness: *proportion of each diagonal term increases within corresponding row and column*

Alternative indexes of assortativeness

- Correlation between spouses matching trait
- Minimum distance (Shen 2020, Fernandez and Rogerson): matching table is linear combination between those obtained under random matching and perfect sorting
- Likelihood ratio (Eika et al. 2019)
  - Simple: ratio of empirical to random matching probabilities
  - Weighted: weighted sum of likelihood ratios
- All but the simple likelihood ratio measure satisfy the minimum conditions – *monotonicity* and *perfect assortativeness*

Data: Current Population Survey

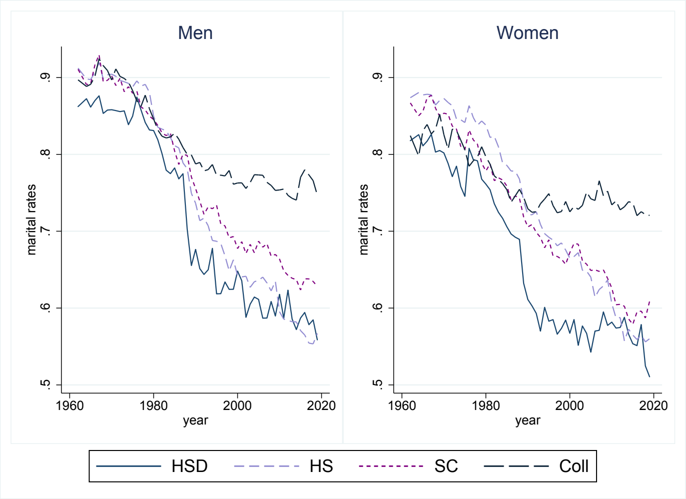
- Annual March waves from 1962 to 2019
- Survey of HH: records information for all adult members interviewed and relationship betwwen them
- Detailed demographic and socio-economic information, including marital status, age, sex, education and earned income
- 4 education groups: high-school dropout (HSD), high-school graduate (HS), some college (SC) and 4+ years College degree (C)
- We will compared 10-year cohorts, from those born in the 30s through the 70s

Marriage

- Self reported marital status does not separate between legal marriage and cohabitation
- Divorced, widows and those never married are counted as singles
- Consider individuals aged 35-44, when marital rates are stable
- For couples we count families where at least one of the spouses is in this age range

Marital rates declined sharply, especially for the least educated

Marital rates, 35-44 year old



Sorting patterns conditional on marriage, birth cohorts 1930-39 and 1970-79

	1930-39				1970-79			
M\W	NQ	Sec	HS	Coll	NQ	Sec	HS	Coll
HSD	18.4	11.3	1.1	0.4	4.5	2.4	1.2	0.5
HS	7.5	24.1	3.3	1.3	2.0	13.5	7.8	5.1
SC	1.4	7.0	3.1	1.4	0.6	4.9	11.5	8.3
C	0.6	6.2	5.1	7.8	0.2	2.5	6.5	28.4

Changes in sorting in a $k \times k$ matching market

- Look at all 2×2 sub-tables
  - Complete map may show assortativeness changing at different directions in different parts of the global matching matrix
- Merged tables: marry one's own type vs marrying anyone else
  - Under the EV assumption, merged index has similar structural interpretation
  - Can include or exclude singles in the definition of 'everyone else'

Changes in sorting in the 4×4 matching market

man \ woman	1	2	3	4
1	s_{11}	s_{12}	s_{13}	s_{14}
2	s_{21}	s_{22}	s_{13}	s_{24}
3	s_{31}	s_{32}	s_{33}	s_{14}
4	s_{41}	s_{42}	s_{43}	s_{44}

- All 2×2 submatrices along the main diagonal

s_{II}	s_{IJ}
s_{JI}	s_{JJ}

for $(I, J) = (1,2), (2,3), (3,4), (1,3), (2,4), (1,4)$

- Merge matrices

s_{II}	$\sum_{J \neq I} s_{IJ}$
$\sum_{J \neq I} s_{JI}$	$\sum_{J,K \neq I} s_{JK}$

for $I = 1, \ldots, 4$.

- Test changes in assortativeness in each 2×2 submatrices using multiple testing step-down procedure of Romano-Wolf

Changes in sorting conditional on marriage: test results

	GEV	SEV	χ^2	MD	ll ratio	weighted sum
			Sub-tables			
HSD vs HS	$\times$					
HS vs SC	$\times$					
SC vs C	$\times$					
HSD vs SC	$\times$					
HS vs C	$\times$					
HSD vs C	$\times$					
			Merged tables excl. education			
HSD	$\times$					
HS	$\times$					
SC	$\times$					
C	$\times$					
			Merged tables incl. education			
HSD	$\times$					
HS	$\times$					
SC	$\times$					
C	$\times$					

Changes in sorting conditional on marriage: test results

	GEV	SEV	χ^2	MD	ll ratio	weighted sum
			Sub-tables			
HSD vs HS	$\times$	0.9^{***}				
HS vs SC	$\times$	0.3^{***}				
SC vs C	$\times$	0.6^{***}				
HSD vs SC	$\times$	0.6^{***}				
HS vs C	$\times$	0.2^{***}				
HSD vs C	$\times$	0.5^{***}				
			Merged tables excl. education			
HSD	$\times$	1.3^{***}				
HS	$\times$	-0.6^{***}				
SC	$\times$	0.4^{***}				
C	$\times$	-0.0				
			Merged tables incl. education			
HSD	$\times$	0.8^{***}				
HS	$\times$	-1.0^{***}				
SC	$\times$	0.1^{***}				
C	$\times$	0.3^{***}				

Changes in sorting conditional on marriage: test results

	GEV	SEV	χ^2	MD	ll ratio	weighted sum
				Sub-tables		
HSD vs HS	$\times$	0.9^{***}	0.14^{***}	0.12^{***}	0.8^{***}	0.15^{***}
HS vs SC	$\times$	0.3^{***}	0.06^{***}	0.10^{***}	0.2^{***}	0.11^{***}
SC vs C	$\times$	0.6^{***}	0.10^{***}	0.02	0.3^{***}	0.14^{***}
HSD vs SC	$\times$	0.6^{***}	0.16^{***}	0.14^{***}	1.6^{***}	0.12^{***}
HS vs C	$\times$	0.2^{***}	0.11^{***}	0.04^{***}	1.0^{***}	0.09^{***}
HSD vs C	$\times$	0.5^{***}	-0.01	0.00	5.0^{***}	-0.01
			Merged tables excl. singles			
HSD	$\times$	1.3^{***}	0.07^{***}	0.08^{***}	4.1^{***}	0.08^{***}
HS	$\times$	-0.6^{***}	-0.08^{***}	-0.18^{***}	-0.0	0.14^{***}
SC	$\times$	0.4^{***}	0.03^{***}	0.13^{***}	0.1^{***}	0.11^{***}
C	$\times$	-0.0	0.01^{*}	0.11^{***}	-0.3^{***}	0.02^{*}
			Merged tables incl. singles			
HSD	$\times$	0.8^{***}	-0.02^{**}	-0.02^{*}	3.3^{***}	-0.02^{*}
HS	$\times$	-1.0^{***}	-0.11^{***}	-0.28^{***}	-0.3^{***}	-0.24^{***}
SC	$\times$	0.1^{***}	0.01^{***}	0.05^{***}	-0.0	0.05^{***}
C	$\times$	0.3^{***}	0.03^{***}	0.17^{***}	-0.1^{***}	0.09^{***}

Counterfactual simulations

- We say: two marriage markets are equally assortative iff sorting is generated by the same structural matrix of marital surplus
  - Does not imply that the two sorting matrices look the same!
- Under SEV
  - can calculate surplus matrix from sorting patterns
  - **can also calculate sorting patterns from surplus matrix**
- Use SEV to determine counterfactual sorting that emulates assortativeness of one market under marginals of another

Counterfactual simulation: superimpose 30's preferences on 70's distribution of education

Rate of singlehood by education

	1970's cohort		Counterfactual using 1930's preferences	
	men	women	men	women
High school dropout	27.58	31.27	12.39	6.57
High school	28.80	30.23	15.18	3.95
Some college	25.75	28.91	16.44	13.71
4+years college	16.40	18.48	6.18	22.67
all	23.62	25.39	12.16	14.17

Counterfactual simulation: superimpose 30's preferences on 70's distribution of education

Marital sorting conditional on marriage

	1970's cohort				Counterfactual using 1930's preferences			
	HSD	HS	SC	C	HSD	HS	SC	C
HSD	4.50	2.44	1.17	0.46	4.02	3.31	1.12	0.56
HS	1.96	13.53	7.79	5.11	3.42	14.68	7.17	4.13
SC	0.63	4.90	11.55	8.28	0.98	6.57	10.25	7.01
C	0.23	2.51	6.52	28.42	0.23	3.41	9.96	23.16

Conclusions

- Defining PAM is easy, but measuring it (or comparing it across populations) is much more challenging
- Here we propose a new index of assortativeness that is explicitly linked to the underlying forces driving it
- The index does not require the structure to give an indication of assortativeness: it is a just a measure of the odds of marrying your like vs marrying someone different
- Contrary to other measure of assortativeness, for the SEV we know the conditions under which changes in the index reflect changes in the value of different matches
- This property is extremely useful for counterfactual analysis: can use SEV to assess extent to which changes in assortativeness are responsible for changes in inequality

Where to find more details

Chiappori, Costa-Dias and Meghir, 2020, ‘Changes in Assortative Matching: Theory and Evidence for the US,’ ▸ https://www.nber.org/papers/w26932

Chiappori, Costa-Dias, Crossman and Meghir, 2020, ‘Changes in Assortative Matching and Inequality in Income: Evidence for the UK,’ ▸ https://onlinelibrary.wiley.com/doi/10.1111/1475-5890.12217

Thank you!