Partial Identification in Matching models for the marriage market

C. Gualdani & S. Sinha - JPE, 2022

Sergi Barcons Llorach, Maria Frech, Iván Rendo Barreiro

Toulouse School of Economics

February 20, 2023

Motivation

- **Has positive educational sorting in the marriage market increased over time?**
  - Driver of intergenerational inequality
- Previous parametric approaches say YES
- Non-parametric approach (this paper) says NOT CLEAR

Introduction

- **One-to-one matching model** with perfectly transferable utilities
  - Interest in recovering the surplus of the match
  - Strong parametric assumptions on the agent's unobserved taste shock

Introduction

- **One-to-one matching model** with perfectly transferable utilities
  - Interest in recovering the surplus of the match
  - Strong parametric assumptions on the agent's unobserved taste shock
- **Methodological contribution**:
  1. Without parametric assumptions severe under-identification
  2. Partial identification under non-parametric assumptions
  3. Reduction of computational burden

Introduction

- **One-to-one matching model** with perfectly transferable utilities
  - Interest in recovering the surplus of the match
  - Strong parametric assumptions on the agent's unobserved taste shock
- **Methodological contribution**:
  1. Without parametric assumptions severe under-identification
  2. Partial identification under non-parametric assumptions
  3. Reduction of computational burden
- **Empirical contribution**
  - Re-examine positive educational matching in the data
  - Previous results driven by parametric assumption

One-to-one Matching Model

One-to-one matching model

with perfectly transferable utilities

A1 One large market framework

- Two-sided market: one side men, $\mathcal{I}$, other side women, $\mathcal{J}$

A2 Finite number of observed types (0 means no match)

- Type of men, $\mathcal{X}$, type of women, $\mathcal{Y}$

A3 Taste shocks

- ϵ_{iy} idiosyncratic preference of man i to marry woman y
- η_{xj} idiosyncratic preference of women j to marry men x

A4 Separability

- Match surplus from remaining single is normalized to zero
- A match between $i \in \mathcal{I}$ of type $x \in \mathcal{X}$ and a woman $j \in \mathcal{J}$ of type $y \in \mathcal{Y}$ generates a match surplus

$$\widetilde{\phi_{ij}} = \phi_{xy} + \epsilon_{iy} + \eta_{xj}$$

Stable Matching

A matching consist of

(i) A measure on the set $\mathcal{I}x\mathcal{J}$ (match assignment)

(ii) A set of payoffs $\{U_i\}_{i\in\mathcal{I}}$ and $\{V_j\}_{j\in\mathcal{J}}$ such that $U_i + V_j = \widetilde{\Phi_{ij}}$ (sharing rule)

Stable Matching

A matching consist of

(i) A measure on the set $\mathcal{I}x\mathcal{J}$ (match assignment)

(ii) A set of payoffs $\{U_i\}_{i\in\mathcal{I}}$ and $\{V_j\}_{j\in\mathcal{J}}$ such that $U_i + V_j = \widetilde{\Phi_{ij}}$ (sharing rule)

Stable matching

A matching is stable when no agent has an incentive to change their partner, i.e.

$$U_i \geq \widetilde{\Phi_{i0}} \quad \forall i \in \mathcal{I}$$
$$V_j \geq \widetilde{\Phi_{j0}} \quad \forall j \in \mathcal{J}$$
$$U_i + V_j \geq \widetilde{\Phi_{ij}} \quad \forall (i,j) \in \mathcal{I}x\mathcal{J}$$

$\Rightarrow$ Equilibrium mass of couples is unique for every (x, y).

Identification

Identification: Parameter(s) of Interest and Data

Parameter of interest

The parameters of interest are **total surplus** Φ_{xy} and U_{xy}, the part of the surplus gained by a man of type x when matching a y-type woman.

$$\Phi_{xy} = U_{xy} + V_{xy}$$

Identification: Parameter(s) of Interest and Data

Parameter of interest

The parameters of interest are **total surplus** Φ_{xy} and U_{xy}, the part of the surplus gained by a man of type x when matching a y-type woman.

$$\Phi_{xy} = U_{xy} + V_{xy}$$

Data

$$p_{y|x} \triangleq \frac{\mu_{xy}}{m_x} \equiv \frac{\#\,\mathrm{couples}\,(\mathrm{x},\mathrm{y})}{\#\,\mathrm{people\,of\,type\,x}}; \quad p_x \triangleq \frac{m_x}{\sum_{x\in\mathcal{X}} m_x}; \quad p_{x|y},\ p_y \quad \text{analogous}$$

Under-identification Without Parametric Assumptions

Proposition. For any data $p_{y|x}$ and $p_{x|y}$, and (true) systematic surplus Φ, there exist (U, V), $\{F_x\}_x$ and $\{G_y\}_y$ (cdfs of taste shocks) such that

$$\begin{aligned} p_{y|x} &= \kappa(U, F_x, y) \\ p_{x|y} &= \kappa(V, G_y, y) \qquad \forall\, x, y \text{ types of people (e.g. education)} \\ U_{xy} + V_{xy} &= \Phi_{xy} \end{aligned}$$

where $\kappa(U, F_x, y)$ is the *model-implied* prob. of marrying a woman y cond. on being a man x.

Under-identification Without Parametric Assumptions

Proposition. For any data $p_{y|x}$ and $p_{x|y}$, and (true) systematic surplus Φ, there exist (U, V), $\{F_x\}_x$ and $\{G_y\}_y$ (cdfs of taste shocks) such that

$$
\begin{aligned}
p_{y|x} &= \kappa(U, F_x, y) \\
p_{x|y} &= \kappa(V, G_y, y) \qquad \forall\, x, y \text{ types of people (e.g. education)} \\
U_{xy} + V_{xy} &= \Phi_{xy}
\end{aligned}
$$

where $\kappa(U, F_x, y)$ is the *model-implied* prob. of marrying a woman y cond. on being a man x.

We need to add distributional assumptions (on unobserved heterogeneity) to obtain identification!

Partial Identification under Non-parametric Assumptions

How to construct the identified set?

(i) Focus on identifying U, then V and Φ from them. Thus, the identified set (for U) is

$$\mathcal{U}^* \equiv \{U \quad \textit{s.t.} \text{ there exist a conditional CDF in a restricted space}$$
$$\text{that rationalizes the model-implied probability } \kappa\}$$

Partial Identification under Non-parametric Assumptions

How to construct the identified set?

(i) Focus on identifying U, then V and Φ from them. Thus, the identified set (for U) is

$$\mathcal{U}^* \equiv \{U \quad s.t. \text{ there exist a conditional CDF in a restricted space} \\ \text{that rationalizes the model-implied probability } \kappa\}$$

Checking if $U \in \mathcal{U}^*$ is a problem that can be transform into a linear program.

Partial Identification under Non-parametric Assumptions

How to construct the identified set?

(i) Focus on identifying U, then V and Φ from them. Thus, the identified set (for U) is

$$\mathcal{U}^* \equiv \{U \quad s.t. \text{ there exist a conditional CDF in a restricted space}$$
$$\text{that rationalizes the model-implied probability } \kappa\}$$

Checking if $U \in \mathcal{U}^*$ is a problem that can be transform into a linear program.

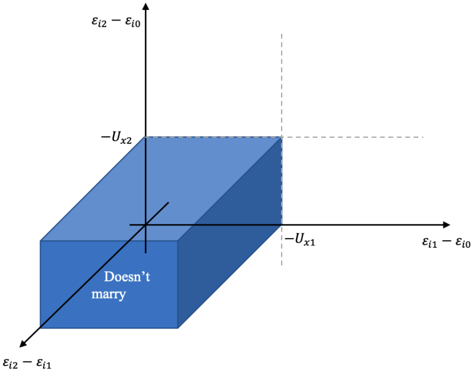
For instance, with 2 types ($\mathcal{X} = \mathcal{Y} = \{1, 2\}$) one of the conditions to check is

$$\exists \{\Delta F_x\}_{x \in \mathcal{X}}, \; s.t. \; p_{0|x} = \Delta F_x(-U_{x1}, -U_{x2}, +\infty)$$

Note: ΔF_x *denotes the distribution of* $(\epsilon_{i1} - \epsilon_{i0}, \epsilon_{i2} - \epsilon_{i1}, ...)$

Partial Identification under Non-parametric Assumptions

How to construct the identified set?



Intuition: man *i* remains single if taste shock for both types of women isn't large enough.

➡ Same logic for the probability of marrying a women of type $y \in \{1, 2\}$ other equations

Only solve the problem for x which belong to a *finite* number of 3-tuples.

$\implies$ can solve finite-dim. problem & extend to a proper family of CDFs.

Partial Identification under Non-parametric Assumptions

Linear programming: without and with non-parametric restrictions

Three different sets of constraints on $\hat{F}$

Partial Identification under Non-parametric Assumptions

Linear programming: without and with non-parametric restrictions

Three different sets of constraints on $\hat{F}$

1. $\Delta\hat{F}_x$ implied probabilities must **match the data**:
   - $p_{0|x} = \Delta\hat{F}_x(-U_{x1}, -U_{x2}, +\infty)$
   - same for $p_{1|x}$ and $p_{2|x}$

Partial Identification under Non-parametric Assumptions

Linear programming: without and with non-parametric restrictions

Three different sets of constraints on $\hat{F}$

1. $\Delta\hat{F}_x$ implied probabilities must **match the data**:
   - $p_{0|x} = \Delta\hat{F}_x(-U_{x1}, -U_{x2}, +\infty)$
   - same for $p_{1|x}$ and $p_{2|x}$
2. $\Delta\hat{F}_x$ must satisfy defining properties of **CDFs**:
   - $\Delta\hat{F}_x(\infty, \infty, \infty) = 1$
   - $\Delta\hat{F}_x(a_1, a_2, -\infty) = 0$
   - some more

Partial Identification under Non-parametric Assumptions

Linear programming: without and with non-parametric restrictions

Three different sets of constraints on $\hat{F}$

1. $\Delta\hat{F}_x$ implied probabilities must **match the data**:
   - $p_{0|x} = \Delta\hat{F}_x(-U_{x1}, -U_{x2}, +\infty)$
   - same for $p_{1|x}$ and $p_{2|x}$
2. $\Delta\hat{F}_x$ must satisfy defining properties of **CDFs**:
   - $\Delta\hat{F}_x(\infty, \infty, \infty) = 1$
   - $\Delta\hat{F}_x(a_1, a_2, -\infty) = 0$
   - some more
3. Those coming from the (non-parametric) **additional assumption** added to recover identification. Here, for instance, $\Delta\hat{F}_x$ is symmetric conditional on x:
   - $\Delta\hat{F}_x(-U_{x1}, \infty, \infty) = 1 - \Delta\hat{F}_x(-U_{x1}, \infty, \infty)$
   - $\Delta\hat{F}_x(0, \infty, \infty) = 1/2$
   - other analogous ones

Reducing Computational Burden

Simplifying grid search

- Need to solve the previous linear program for every point of a grid for $\mathcal{U}$.
- The complexity of the problem increases *exponentially* with the number of types.

Simplifying grid search

- Need to solve the previous linear program for every point of a grid for $\mathcal{U}$.
- The complexity of the problem increases *exponentially* with the number of types.

Solution: reduce the problem to only solve for some $U \in \mathcal{U}$. Summary of procedure:

1. generate a grid for $\mathcal{U}$
2. fix an order for the 3-tuples described before, only when the position of *critical pairs* is different, the associated values of U can give different solutions
   - ➥ *formally find the π-ordering of each U of the grid*
3. collect those producing the same π-ordering in equivalence classes
4. select a single representative of each equiv. class and solve the linear program

Empirical Application

Empirical Application: Educational Sorting

Positive educational sorting driven mechanically by changes in shares of educated people?

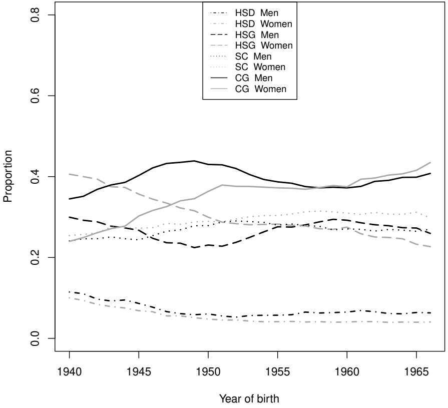
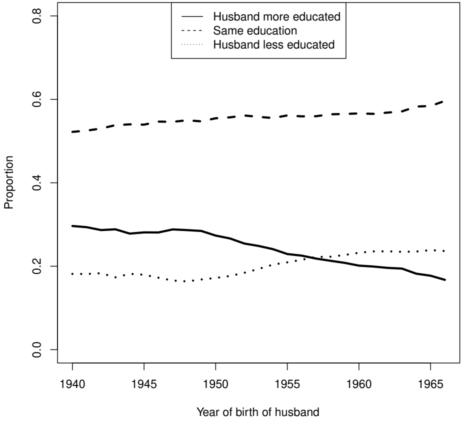


Figure 1: Education of men and women.



(A)

Empirical Application: Educational Sorting

- Data:
  - American Community Survey
  - 4 educational levels
  - 28 cohorts
  - first marriages and never-married singles
- Previous empirical approaches find **positive educational sorting** has **increased** in the U.S.
- This paper suggest this result is driven by parametric Logit assumption.

Empirical Application: Educational Sorting

Estimated identified sets of surplus Φ

Assumptions on unobservables	Wife $\rightarrow$	1	2	3	4
	Husband $\downarrow$		Early cohorts		
Logit		-2.06	-3.07	-5.22	-8.55
[A]	1	-2.06	$[-3.07, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
[B]		-2.06	$[-3.07, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
Logit		-3.73	-1.35	-3.4	-5.76
[A]	2	$[-5.3, -3.48]$	$[-3.79, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
[B]		$[-5.3, -3.48]$	$[-3.63, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
Logit		-5.29	-2.47	-2.12	-4.32
[A]	3	$(-\infty, -4.69]$	$(-\infty, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
[B]		$(-\infty, -4.69]$	$(-\infty, +\infty)$	$(-\infty, +\infty)$	$(-\infty, +\infty)$
Logit		-8.01	-4	-2.46	-1.11
[A]	4	$(-\infty, -6.59]$	$(-\infty, +\infty)$	$(-\infty, +\infty)$	$[-5.74, +\infty)$
[B]		$(-\infty, -6.59]$	$(-\infty, +\infty)$	$(-\infty, +\infty)$	$[-4.4, +\infty)$

Details **[A]** & **[B]** & Estimated identified sets & Logit estimates

Conclusion

- Previous empirical matching models no information about Φ without parametric assumptions
- Propose computational approach for constructing identified set of Φ
- Use methodology to re-examine empirical literature on marriage market
    - Results inconclusive about evolution of educational sorting

Appendix

Definition of an element in the identified set

$$U \in \mathcal{U}^* \iff \exists \{\Delta F_x\}_{x \in \mathcal{X}} \in \Delta F^\dagger \; s.t. \; p_{y|x} = \kappa(U, \delta F_x, y) \;\; \forall (x, y) \in \mathcal{X} \times \mathcal{Y}_0$$

In the case of $\mathcal{X} = \mathcal{Y} = \{1, 2\}$ and assuming $\Delta F^\dagger = \Delta F$, it reduces to

$$\begin{aligned}
&\exists \{\Delta F_x\}_{x \in \mathcal{X}} \in \Delta F \; s.t. \; \forall x \in \mathcal{X}, \\
&p_{1|x} = 1 + \Delta F_x(-U_{x1}, \infty, U_{x2} - U_{x1}) - \Delta F_x(\infty, \infty, U_{x2} - U_{x1}) - \Delta F_x(-U_{x1}, \infty, \infty) \\
&p_{2|x} = \Delta F_x(\infty, \infty, U_{x2} - U_{x1}) - \Delta F_x(\infty, -U_{x2}, U_{x2} - U_{x1}) \\
&p_{0|x} = \Delta F_x(-U_{x1}, -U_{x2}, \infty)
\end{aligned}$$

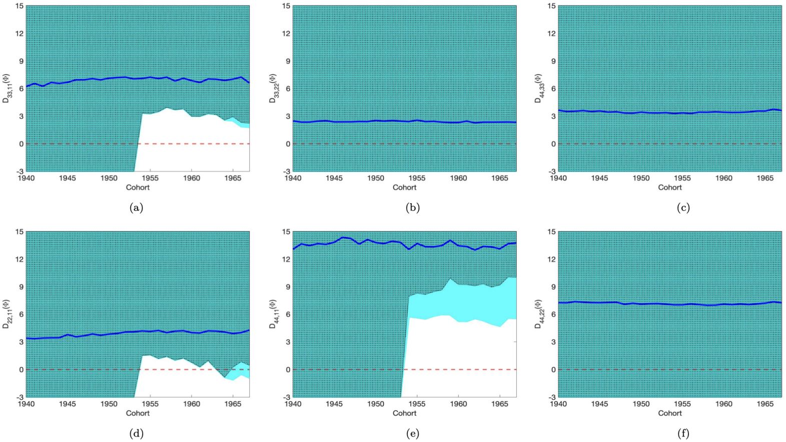
back

Nonparametric assumptions in $\{\Delta F_x\}_{x \in \mathcal{X}}$

1. $\Delta\epsilon_i$ is independent of X_i.
2. Conditional on X_i, and for each $l \in \{1, \dots, d\}$, $\Delta\epsilon_{i,l}$ has a distribution symmetric at 0.
3. Conditional on X_i, $\{\Delta\epsilon_{i,l}\}_{l \in \{1,\dots,d\}}$ are identically distributed.
4. Conditional on X_i, $\{\Delta\epsilon_i^y\}_{y \in \mathcal{Y}_0}$ are identically distributed.

Specification **[A]** imposes 4 and **[B]** imposes 2, 3 and 4. back

Estimated identified sets and Logit estimates



where $D_{xx,\widetilde{x}\widetilde{x}}$ = incremental value of marrying a more educated man $(x > \widetilde{x})$ as the education of the woman increases $(x > \widetilde{x})$. back