Exploring the Causes of Changing Marriage Patterns

Pierre-André Chiappori, Bernard Salanié, Avi Tillman and Yoram Weiss

March 4, 2008
E.g. Burtless (EER 1999): over 1979-1996,

*The changing correlation of husband and wife earnings has tended to reinforce the effect of greater pay disparity.*

Maybe 1/3 of the increase in *household-level* inequality (Gini) comes from rise of single-adult households and 1/6 from increased assortative matching.
Changes in Household Size

![Graph showing changes in household size from 1960 to 2006](#)
Changes in Endogamy

Burtless again, US 1979-1996:

*The Spearman rank correlation of husband and wife earnings increased from 0.012 to 01.45.*

i.e: the average absolute difference of percentiles went from 0.406 to 0.377 (zero correlation would give 0.408).

Here we focus on “educational endogamy” in the US, on “educational*origin endogamy” in Israel.

*Education endogamy:*

Can be measured e.g. in a contingency table of education classes $k, l$ by comparing 1 to

$$\frac{N_{kl}N}{N_k N_l}.$$
### Educational Endogamy in the US

#### Year – 1962 – US

<table>
<thead>
<tr>
<th>Men - Women</th>
<th>HSD</th>
<th>HSG</th>
<th>SC</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSD</td>
<td>2.433</td>
<td>1.178</td>
<td>0.483</td>
<td>0.241</td>
</tr>
<tr>
<td>HSG</td>
<td>0.849</td>
<td>1.738</td>
<td>0.945</td>
<td>0.426</td>
</tr>
<tr>
<td>SC</td>
<td>0.362</td>
<td>0.846</td>
<td>1.543</td>
<td>0.713</td>
</tr>
<tr>
<td>CG</td>
<td>0.164</td>
<td>0.331</td>
<td>1.006</td>
<td>2.792</td>
</tr>
</tbody>
</table>

#### Year – 1973 – US

<table>
<thead>
<tr>
<th>Men - Women</th>
<th>HSD</th>
<th>HSG</th>
<th>SC</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSD</td>
<td>3.708</td>
<td>1.505</td>
<td>0.529</td>
<td>0.155</td>
</tr>
<tr>
<td>HSG</td>
<td>1.186</td>
<td>2.169</td>
<td>0.878</td>
<td>0.272</td>
</tr>
<tr>
<td>SC</td>
<td>0.428</td>
<td>0.900</td>
<td>1.406</td>
<td>0.547</td>
</tr>
<tr>
<td>CG</td>
<td>0.106</td>
<td>0.242</td>
<td>0.842</td>
<td>2.368</td>
</tr>
</tbody>
</table>

#### Year – 1983 – US

<table>
<thead>
<tr>
<th>Men - Women</th>
<th>HSD</th>
<th>HSG</th>
<th>SC</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSD</td>
<td>5.554</td>
<td>1.736</td>
<td>0.753</td>
<td>0.166</td>
</tr>
<tr>
<td>HSG</td>
<td>1.879</td>
<td>2.609</td>
<td>0.983</td>
<td>0.239</td>
</tr>
<tr>
<td>SC</td>
<td>0.435</td>
<td>1.044</td>
<td>1.438</td>
<td>0.532</td>
</tr>
<tr>
<td>CG</td>
<td>0.244</td>
<td>0.279</td>
<td>0.704</td>
<td>1.797</td>
</tr>
</tbody>
</table>
Assortative Matching

Becker 1973: in equilibrium on the marriage market, if e.g. education of husband and wife are complements then partners will match within education groups.

Yet women used to marry richer and more educated men, in part because men tended to be richer and more educated → sex ratio;

As women have become more educated and hold jobs, that effect operates less strongly.

But it may also be that women’s preferences (or men’s) have changed → change in preferences.

Can we quantify these effects?
In this paper

We show how to identify/estimate a matching model of the marriage market under simplifying assumptions that turn it into a multilogit model = analysis of variance, with a structural interpretation that allows for welfare analysis.

We use our method to investigate

1. the reasons and extent of changes in the return to education on the marriage market in the US;
2. the causes of increased intermarriage among ethnic/education groups in Israel;
3. (later...) France and other countries.
(Some) Background

= a static, frictionless model of matching with transferable utility.

- **static**: we will in fact let singles plan their age 18 to age 35 choice on the marriage market (but no divorce or mortality)
- **frictionless**: changes in transportation, communications, ... may impact our estimates
- **transferable utility**: reduces the dimensionality of the problem.

There are $M$ men, $W$ women. $i$ is a man, $j$ a woman; 0 is celibate.
Given transferable utility, what matters is the surplus $z_{ij}$ of any possible match.
Efficient, stable matching

First group of results comes from Shapley-Shubik (the assignment game, 1972). Since there are no frictions (and information is perfect) matching maximizes the total surplus

\[
\sum_{i=0}^{M} \sum_{j=0}^{W} a_{ij} z_{ij}
\]

given the feasibility constraints \( \forall i, j, \ a_{ij} \geq 0 \) and

\[
\forall j > 0, \sum_{i} a_{ij} \leq 1
\]

\[
\forall i > 0, \sum_{j} a_{ij} \leq 1.
\]
The Dual Problem

If we want to evaluate returns to education on the marriage market, we need to define the “value” of a marriage prospect.

*It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.*

Pride and Prejudice, *opening sentence.*

Although it may be prejudice more than perfect information:

*However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered as the rightful property of some one or other of their daughters.*

Back to earth: we use the dual problem of the linear programming surplus maximization.
The Dual Problem

\[
\min_{u,v} \left( \sum_{i=1}^{M} u_i + \sum_{j=1}^{W} v_j \right)
\]

given that

\[\forall i,j, \; u_i + v_j \geq z_{ij}.\]

\(u_i\) is the surplus man \(i\) gets in equilibrium/optimum. It is also the “price” (in terms of surplus) that his match will have to forgo to marry him: she gets \(v_j = z_{ij} - u_i = \max_k (z_{kj} - u_k)\).

(and \(u_0 = v_0 = 0\)).
Any assignment that solves the primal (or the dual) is Pareto-efficient by construction.
It is stable in the Gale-Shapley sense: no (man,woman) pair could be a blocking coalition.
In fact it is even in the core of the assignment game.
And the extremal stable matchings are implementable by the deferred acceptance mechanism under perfect information.
Once a year in each village the maidens of age to marry were collected all together into one place; while the men stood round them in a circle. Then a herald called up the damsels one by one, and offered them for sale. He began with the most beautiful. When she was sold for no small sum of money, he offered for sale the one who came next to her in beauty. All of them were sold to be wives. The richest of the Babylonians who wished to wed bid against each other for the loveliest maidens, while the humbler wife-seekers, who were indifferent about beauty, took the more homely damsels with marriage-portions. [...] This was the best of all their customs, but it has now fallen into disuse.
Too many unknowns

There is only \((M + W)\) of the \(u\)'s and \(v\)'s, not \(MW\) of the \(z\)'s... but how are we to infer them from the data?

In practice the data we are likely to get is only (at best)

- (usually discrete) characteristics \(X_i\) and \(X_j\) of \(M\) men and \(W\) women unmarried at the beginning of a period
- and who marries whom during that period.

Given large “cells”, this identifies a “matching function”

\[
\mu(X_i, X_j) = \Pr(a_{ij} = 1|X_i, X_j)
\]

where \(a_{ij}\) is the equilibrium assignment/matching.

(which depends on characteristics unobserved by the econometrician, even with frictionless markets.)
Equilibrium conditions

Equilibrium implies that
i is matched with j iff

\[ u_i = z_{ij} - v_j, \]
\[ u_i \geq z_{ik} - v_k \text{ for all } k, \text{ and } u_i \geq z_{i0}, \]
and
\[ v_j \geq z_{kj} - u_k \text{ for all } k, \text{ and } v_j \geq z_{0j}. \]

This is all we have, and we could indeed just use these inequalities, along with a parametric specification of the \( z_{ij} \), to at least partially identify the model (cf Fox (2007)). We choose (at this stage) to simplify, so as to get full identification + very simple estimation.
Assume that $X = G$: a group variable $G$ (education, ethnic group)
(later we introduce more covariates)

**Assumption S (separability)**

$$z_{ij} = Z(G_i, G_j) + \varepsilon_i(G_i, G_j) + \eta_j(G_i, G_j).$$

So conditional on $(G_i, G_j)$, the variation in the surplus from a match separates additively between the partners.
this will imply that each individual has preferences over groups, not over particular persons in each group.
Theorem

Under assumption (S), there exist functions $U$ and $V$ s.t.
for any matched couple $(i,j)$,

$$u_i = U(G_i, G_j) + \varepsilon_i(G_i, G_j)$$

and

$$v_j = V(G_i, G_j) + \eta_j(G_i, G_j).$$

So $u_i$ and $v_j$ depend

1. on observables of the individual and partner
2. on an error term that does not depend on unobserved characteristics of the partner, conditional on his or her group.
A Bit More Useful

Necessary and sufficient conditions for men to be stable

1. For all matched couples \((i \in G, j \in H)\),

\[
\varepsilon_i(G, H) - \varepsilon_i(G, K) \geq U(G, K) - U(G, H) \quad \text{for all } K \tag{2}
\]

\[
\varepsilon_i(G, H) - \varepsilon_i(G, 0) \geq U(G, 0) - U(G, H); \quad \tag{3}
\]

2. For all single males \(i \in G\),

\[
\varepsilon_i(G, K) - \varepsilon_i(G, 0) \leq U(G, 0) - U(G, K) \quad \text{for all } K. \tag{4}
\]

These are just the defining equations for a multinomial choice model.
Why not a multilogit then?

Assuming that the $\varepsilon'$ s are type-I extreme value is a natural choice.
it also gives some nice formulæ (already in Choo and Siow *JPE* 2006):
Normalizing $U(G, 0) = V(0, H) = 0$,

$$\log \frac{U(G, H) + V(G, H)}{2} = \frac{|(G,H) \text{ marriages}|}{\sqrt{|\text{at risk in } G| \times |\text{at risk in } H|}}.$$ 

and if $i \in G$ and $j \in H$ match,

$$u_i = \log \frac{|(G,H) \text{ marriages}|}{|\text{at risk in } G|}.$$
We can only define “large” groups for estimation, but we also have other covariates: cohort and age will be used here income could also be used. So now $X = (G, p)$ where $p$ are other personal covariates, and everything generalizes to

$$z_{ij} = Z(G_i, G_j) + \varepsilon_i(X_i, G_j) + \eta_j(G_i, X_j).$$

(maintaining additive separability.)
In practice

We rely on datasets (Labor Force Surveys in Israel, June CPS in the US) that give us information on:

- group variables $G$ of individuals and their partner, if any
- their date of marriage (first? current? a difficulty here—so we discard men and women older than 30 in Israel, 35 in the US)
- some other covariates (we use the age of the individual.)

We reconstitute in every year a population of unmarried men and women and we estimate the model on marriage patterns.
**Specification**

When a man reaches age 18, he plans when and whom to marry.
If man $i$ born in cohort $c_i$ marries woman $j$ at age $a_j$, his (mean) utility is

$$U_{ij} = \alpha_1(G_i, G_j, a_i) + \alpha_2(G_i, G_j, a_i, c_i) + \text{spline}(c_i, \alpha(G_i, G_j)) + \varepsilon_{ij}.$$  

We assume that $\varepsilon_{ji}$ is type I-EV with standard variance; we normalize $U_{i0} = U_{j0} = 0$.

- $\alpha_1(G_i, G_j, a_i)$ (quadratic in $a_i$) accounts for age-dependent but cohort-independent group preferences;
- $\alpha_2(G_i, G_j, a_i, c_i)$ (includes $a_i \ast c_i$, $a_i^2 \ast c_i$ and $a_i \ast c_i^2$) allows for changes in the age profile of group preferences across cohorts;
- $\text{spline}(c_i, \alpha(G_i, G_j))$ is a flexible (6-knots natural cubic spline) specification for cohort-dependent but age-independent group preferences.
Has more preference for assortative matching amplified the increases in wage inequality?

We use the June CPS, 6 waves from 1971 to 1995. We focus on whites (and white-white matches: 95% of matched whites.)

4 educational levels:

- high school dropouts (14,575 M, 10,361 W)
- high school graduates (35,744 M, 36,096 W)
- some college (21,877 M, 21,199 W)
- college graduates (19,134 M, 17,400 W).
Women High School Dropouts: Surplus from Marrying
Women High School Dropouts: Surplus from Exogamy

Age at marriage
Mean utility over endogamy
−5
0
20 25 30 35
High School Dropouts
High School Graduates
Some College
College Graduates
Born 1945
Born 1955
Born 1965
Women College Graduates: Surplus from Marrying

Mean utility over celibacy

Age at marriage

Some College
College Graduates

High School Dropouts
High School Graduates

Born 1945
Born 1955
Born 1965
Women College Graduates: Surplus from Exogamy

Figure: Graph showing the mean utility over endogamy for different education levels and birth years. The x-axis represents the age at marriage, ranging from 20 to 35 years. The y-axis represents the mean utility over endogamy, ranging from -5 to 0. The graph compares High School Dropouts, High School Graduates, Some College, and College Graduates across birth years 1945, 1955, and 1965.
Men High School Dropouts: Surplus from Marrying

![Graph showing the mean utility over celibacy for different education levels and birth years.](image)
Men High School Dropouts: Surplus from Exogamy

Age at marriage

Mean utility over endogamy

Some College  College Graduates

High School Dropouts  High School Graduates

Born 1945  Born 1955  Born 1965

●  ●  ●
Men College Graduates: Surplus from Marrying

Mean utility over celibacy

Age at marriage

Born 1945
Born 1955
Born 1965

Some College
College Graduates

High School Dropouts
High School Graduates
Men College Graduates: Surplus from Exogamy

The figure illustrates the mean utility over endogamy against the age of marriage for different educational levels and birth years. The graph shows distinct trends for:

- **Some College**
- **College Graduates**
- **High School Dropouts**
- **High School Graduates**

### Key Insights:

- **Educational Levels**:
  - Some College: The mean utility over endogamy decreases as the age of marriage increases.
  - College Graduates: The mean utility over endogamy remains relatively stable across different ages.
  - High School Dropouts: Similar trend to Some College, decreasing utility with increased age of marriage.
  - High School Graduates: Decreasing trend similar to Some College, but with a slightly different pattern.

### Birth Years**:

- **Born 1945**
- **Born 1955**
- **Born 1965**

The figure suggests that the utility over endogamy decreases with age at marriage, with variations across educational levels and birth years.
Men: Expected Utilities from Marriage

![Graph showing expected utilities for different birth years and educational levels.](image-url)
Women: Expected Utilities from Marriage

![Graph showing expected utilities from marriage for different birth years and education levels.](image)

- Born 1945, HSD
- Born 1955, HSD
- Born 1965, HSD
- Born 1945, CG
- Born 1955, CG
- Born 1965, CG

---

**Expected utilities**

Age

---

**Figure:**
Summary on the US

The main finding

Returns to education on the marriage market have increased for whites not very much for men spectacularly for women.
Application to Israel

“It is well-known that in Israel the traditional Ashkenazim/Sephardim division has blurred”.

- Ashkenazim = originally from Palestine to Italy to Germany (“Ashkenaz”) and to central and Eastern Europe
- Sephardim (≃ Mizrahim) = originally in Spain
- Russian = former Soviet Union
- Israeli = born in Israel, father born there too.
The Israeli groups

1. no BA, Sephardi
2. no BA, Israeli
3. no BA, immigrant from former USSR
4. no BA, Ashkenazim
5. BA, Sephardi
6. BA, Israeli
7. BA, immigrant from former USSR
8. BA, Ashkenazim.
Analysis of Variance for Israeli Men

Based on relative odds ratio.

Table: ANOVA for men

<table>
<thead>
<tr>
<th>Source</th>
<th>Partial SS</th>
<th>df</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>30265.18</td>
<td>615</td>
<td>11.30</td>
</tr>
<tr>
<td>$G_i$</td>
<td>1815.46</td>
<td>7</td>
<td>59.55</td>
</tr>
<tr>
<td>$c_i$</td>
<td>1151.17</td>
<td>34</td>
<td>7.77</td>
</tr>
<tr>
<td>$G_j$</td>
<td>1471.36</td>
<td>8</td>
<td>42.23</td>
</tr>
<tr>
<td>$G_i \times G_j$</td>
<td>16428.43</td>
<td>56</td>
<td>293.36</td>
</tr>
<tr>
<td>$G_i \times c_i$</td>
<td>1609.26</td>
<td>238</td>
<td>6.76</td>
</tr>
<tr>
<td>$G_j \times c_i$</td>
<td>1948.66</td>
<td>272</td>
<td>7.16</td>
</tr>
<tr>
<td>Residual</td>
<td>6497.97</td>
<td></td>
<td>$R^2 = 0.8232$</td>
</tr>
</tbody>
</table>
Different groups have markedly different preferences, and these differences are very stable over time.
Use FFS (GGS?) data in Europe, more recent data in US if possible
Introduce “interactions” in individual preferences
Less restrictive assumptions on errors (nested logit?)