Assortative Matching on the Marriage Market: A Structural Investigation

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

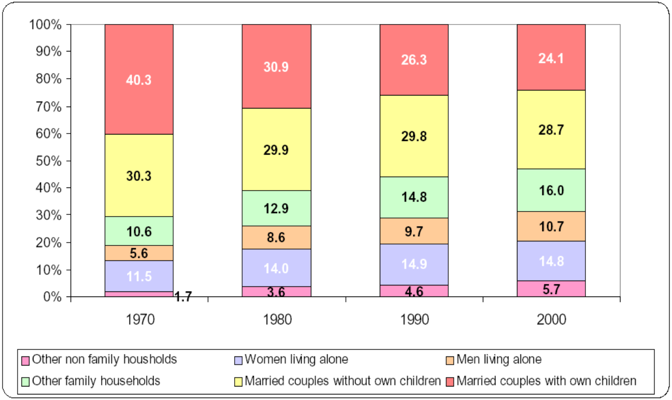
November 12, 2008

Changes in Marriage Patterns since the 60s

- fewer marriages
- more single-person households
- less disparity in education of partners (or an inversion?)
- and also in participation in workforce
- increasing correlation of husband-wife earnings.

Changes in Household Types

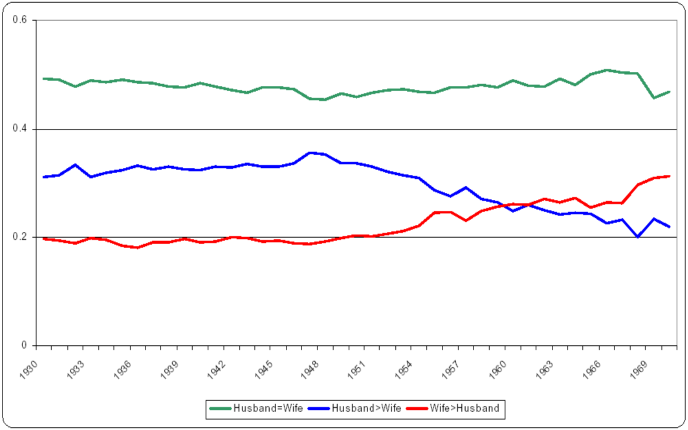
Figure 4a: Households by Type: Selected Years, US



Source: US Census.

Wives' Education

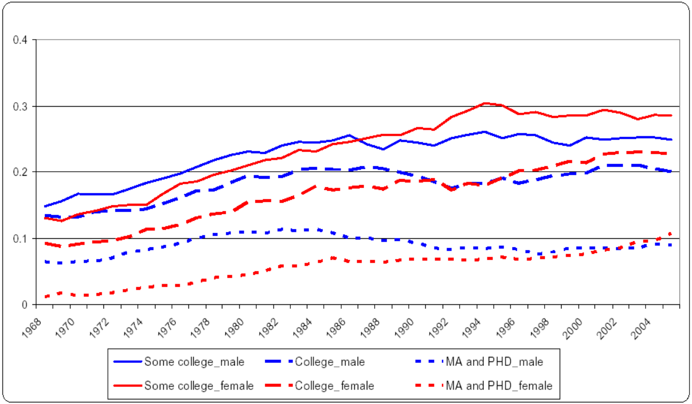
Figure 14: Education of Spouses, by Husband's Year of Birth, US



Source: Current Population Surveys.

Education of US Women

Figure 13: Completed Education by Sex, Age 30-40, US 1968-2005



Source: Current Population Surveys.

Marriage and Inequality in the US

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.

Correlation in UK Couples

Figure 2: Correlation in earnings across husbands and wives. Own calculations from the FES.

Changes in Endogamy in the US

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 in another paper.

Basic Question

Can we attribute *all* the changes in earnings correlation to a (mechanical) effect of increased female education/participation?
Or have preferences over assortative matching changed?

This Paper

- Provides a method that can be used to empirically disentangle these effects
- Based on matching theory
- Leads to multinomial discrete choice models
- Provides a structural interpretation of the results.

Marriage as Matching

Key idea: marriage market as a static matching model (Becker 1974)

- Marriage as a frictionless matching game
- Trait: income, wage, education, ...
- Each marriage generates some surplus
- Framework: transferable utility $\longrightarrow$ the surplus is divided between spouses.

Assortative Matching

Consequences:

1. existence and uniqueness of the stable match (maximizes total surplus)
2. if surplus supermodular (because of household public goods), then positive assortative matching
3. stability requirements generate constraints on intrahousehold allocations (BCW 2004, IW 2006, CIW 2006, 2007, 2008,. . .)

(Some) Assumptions

We rely on 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.

The Model

There are M men, W women. i is a man, j a woman;
"married to 0" ≡ single.
Given transferable utility, what matters is the surplus z_{ij} of any possible match (i, j)
A match is stable if:

- there are no men or women (married or single) who wish to form a new union
- there are no men or women who are married but wish to be single.

Basic Result (Shapley-Shubik 1972)

A match is stable if and only if it maximizes total surplus.
Intuition: duality in linear programming. Primal problem is:

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

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

$$\begin{aligned} \forall j > 0,\ \sum_i a_{ij} &\leq 1 \\ \forall i > 0,\ \sum_j a_{ij} &\leq 1. \end{aligned}$$

The Dual Problem

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

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 $u_0 = v_0 = 0$).

Efficiency

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.

Herodotus, Gale-Shapley, and the “wisest custom” of the Babylonians

...Censored...

History, *I, 1, 196.*

Taking the Model to Data

The primitives of the problem are the *MW* values of the z's; they determine the u's and v's and thus the answers to questions about changes in welfare, returns to education and so on.
How are we to infer them from the data?

What is 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 an "equilibrium matching function"

$$\mu^E(X_i, X_j) = \Pr(a_{ij}^E = 1 | X_i, X_j).$$

Problem: the matching function describes the equilibrium assignment/matching a_{ij}^E,
which depends on the whole matrix of actual z's.
But we cannot hope to recover this huge z matrix.

Equilibrium conditions

Equilibrium implies that
i is matched with j iff

$$
\begin{aligned}
u_i &= z_{ij} - v_j, \\
u_i &\geq z_{ik} - v_k \text{ for all } k\text{, and } u_i \geq z_{i0}, \\
&\quad \text{and} \\
v_j &\geq z_{kj} - u_k \text{ for all } k\text{, and } v_j \geq z_{0j}.
\end{aligned}
$$

The Basic Identification Assumption

Assume that $X = G$: a group variable G (education, ethnic group)
(later we introduce more covariates);
Define $Z(G, H) = E(z_{ij}|G_i = G, H_i = H)$ the average surplus from a match of a G-man and a H-woman.
We want to recover changes in $Z(G, H)$ over time.
Denote $\nu_{ij} = z_{ij} - E(z_{kl}|G_k = G_i, G_l = G_j)$.

Assumption S (separability)

If $G_i = G_k$ and $G_j = G_l$, then

$$\nu_{ij} - \nu_{il} = \nu_{kj} - \nu_{kl}.$$

So conditional on (G_i, G_j), the variation in the surplus from a match separates additively between the partners.

What this buys us

Theorem

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

$$u_i = U(G_i, G_j) + \varepsilon_{i,G_j}$$

and

$$v_j = V(G_i, G_j) + \eta_{j,G_i}.$$

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.**

Corollary

Necessary and sufficient conditions for men to be stable

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

$$\varepsilon_{i,H} - \varepsilon_{i,K} \geq U(G,K) - U(G,H) \text{ for all } K \quad (2)$$
$$\varepsilon_{i,H} - \varepsilon_{i,0} \geq U(G,0) - U(G,H); \quad (3)$$

2. for all single males $i \in G$,

$$\varepsilon_{i,K} - \varepsilon_{i,0} \leq U(G,0) - U(G,K) \text{ for all } K. \quad (4)$$

The Strategy

The conditions in the Corollary are just the defining equations for a multinomial choice model.
We need to add “time” (cohort=birth year), and age too; so $X = (G, c, a)$.
All works as well if

$$z_{ij} = Z(G_i, G_j) + \varepsilon_{i,G_j}(c_i, a_i) + \eta_{j,G_i}(c_j, a_j).$$

Specification Issues

What do we need to assume about the U and V functions and the joint distribution of the ε_{iH} and η_{jG}?
ages 18 to 35 $\otimes$ cohorts 1936 to 1977 $\otimes$ 4 groups
= a lot of parameters to estimate.
Moreover, the joint distribution of ε's and η's is a "big" object; and assumptions about it are crucial to answer questions about changes in preferences.

Choo and Siow (*JPE* 2006)

When there are only group covariates standard multilogit gives some nice formulæ
Normalizing $U(G,0) = V(0,H) = 0$,

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

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

$$u_i = \log \frac{|\text{(G,H) marriages}|}{|\text{at risk in } G|}.$$

Equations vs unknowns

Without covariates:

- equations: 16 independent observed probabilities of matches
- unknowns (1): 16 independent $Z(G, H)$ values
- unknowns (2): 80 independent variance-covariance terms.

With covariates:

- equations: $2,236,144$ independent observed probabilities of matches
- unknowns (1): $24,192$ independent $Z(G, H)$ values
- unknowns (2): $64,668,676$ independent variance-covariance terms.

In practice

We rely on datasets (June CPS, 6 waves from 1971 to 1995) 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 older than 35)
- some other covariates (we use the cohort and age of the individual.)

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

Specification

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).

Simple logit:
We assume that errors are iid type I-EV with standard variance; we normalize $U_{i0} = U_{j0} = 0$.
Nested logits: correlation within group of partner, or within age.
For each (G, H), $U(c, a)$ and $V(c, a)$ are specified as

- a quadratic function of age
- an interaction of age and cohort
- a 3-knot cubic spline on cohort alone.

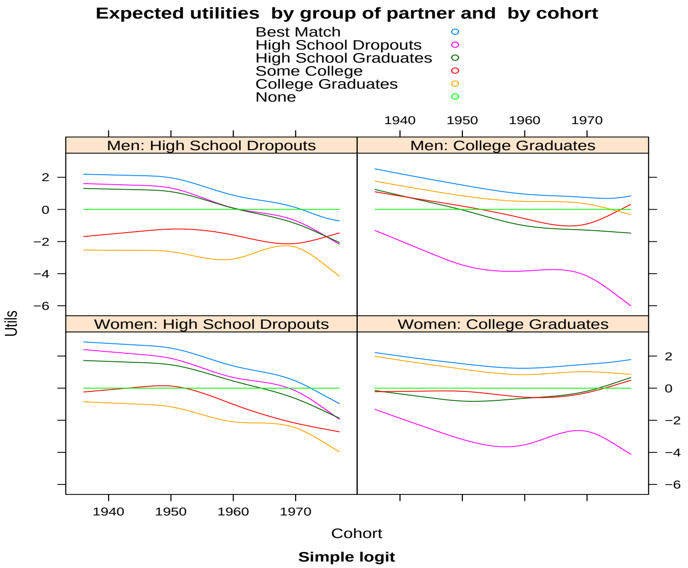
For the coming graphs

We take as our subjects "extreme" groups: high school dropouts and college graduates.
We plot mean utilities from marrying $\max_a U(G, H; c, a)$, idem for V
then mean utilities over endogamy
$\max_a U(G, H; c, a) - \max_a U(G, G; c, a)$
then returns to education on the marriage market

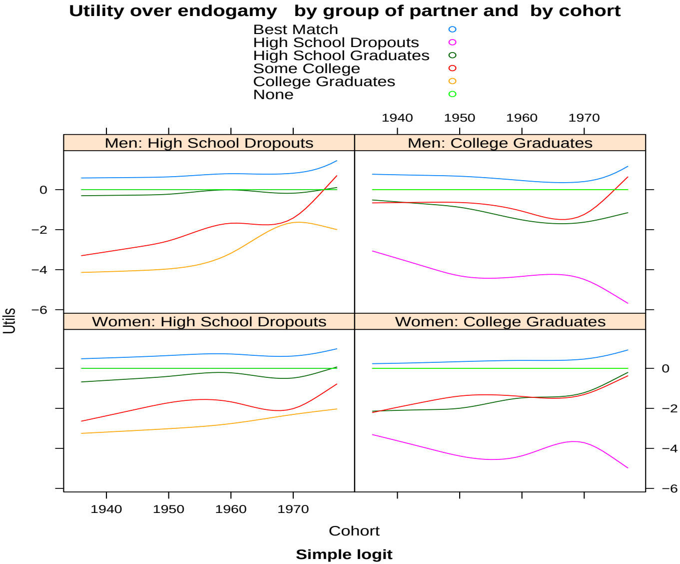
$$\max_{a,H} U(G_1, H; c, a) - \max_{a,H} U(G_2, H; c, a)$$

then mean surpluses from matches $Z(G, H; c)$.

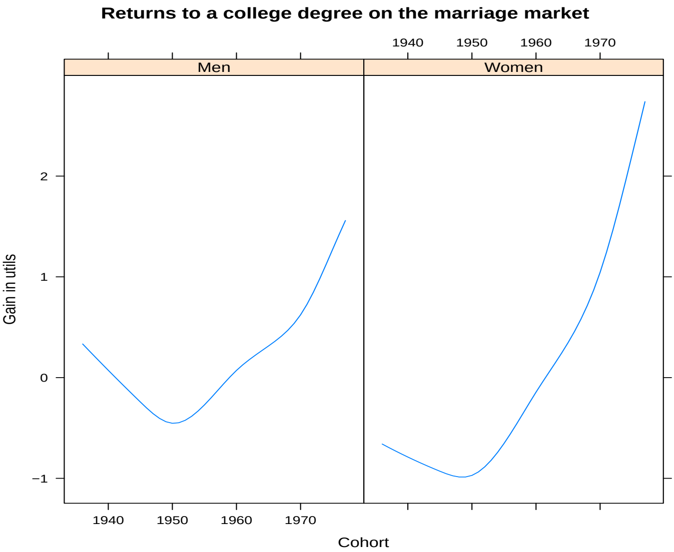
Mean Utility from Marrying



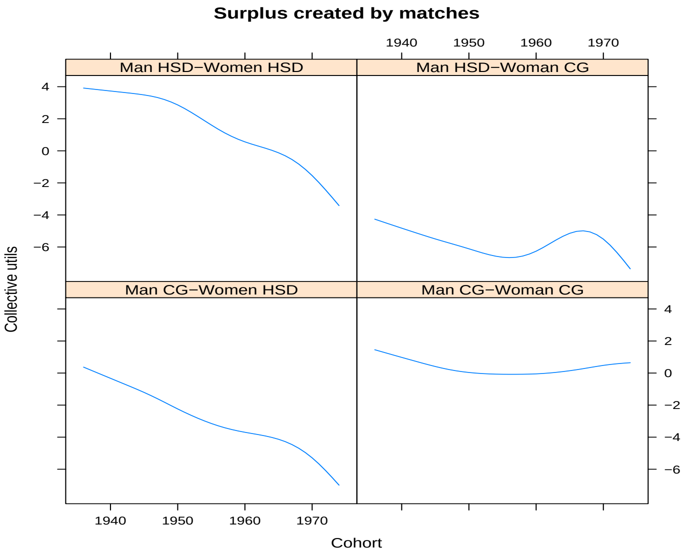
Mean Utility over Endogamy



Returns to Education



Surplus from Matches



Provisional findings

Subject to many caveats on identification (or robustness):

1. surpluses from matches have fallen, except between college graduates
2. returns to education on the marriage market have increased for women, less so for men.