Exploring the Causes of Changing Marriage Patterns

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

March 19, 2008

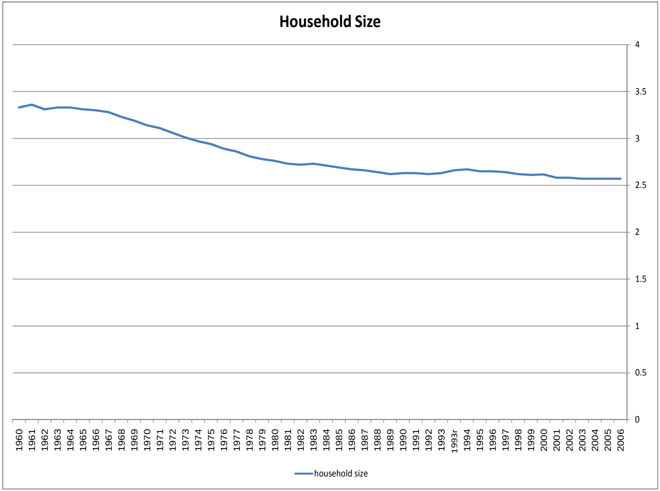
Marriage and Inequality in the US

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.

Changes in Household Size



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

$$\frac{N_{kl}N_{..}}{N_{k.}N_{.l}}.$$

with N_{kl}=number of marriages between classes k and l.

Educational Endogamy in the US

Year – 1962 – US

Men - Women	HSD	HSG	SC	CG
HSD	2.433	1.178	0.483	0.241
HSG	0.849	1.738	0.945	0.426
SC	0.362	0.846	1.543	0.713
CG	0.164	0.331	1.006	2.792

Year – 1973 – US

Men - Women	HSD	HSG	SC	CG
HSD	3.708	1.505	0.529	0.155
HSG	1.186	2.169	0.878	0.272
SC	0.428	0.900	1.406	0.547
CG	0.106	0.242	0.842	2.368

Year – 1983 – US

Men - Women	HSD	HSG	SC	CG
HSD	5.554	1.736	0.753	0.166
HSG	1.879	2.609	0.983	0.239
SC	0.435	1.044	1.438	0.532
CG	0.244	0.279	0.704	1.797

Assortative Matching

Becker 1973: marriage produces collective surplus
if e.g. education of husband and wife are complements in producing surplus, then
in equilibrium on the marriage market, 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
$\longrightarrow$ **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 $\longrightarrow$ **change in preferences.**
Can we disentangle 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 multinomial choice 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

Becker (1973, 1974): *A Theory of Marriage*
= 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;
"married to 0" $\equiv$ single.
Given transferable utility, what matters is the surplus z_{ij} of any possible match.

Efficient, stable matching

A 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

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

The Dual Problem

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

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

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
And the extremal stable matchings are implementable by the deferred acceptance mechanism under perfect information.

Herodotus and the “wisest custom” of the Babylonians

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.

History, *I, 1, 196.*

Too many unknowns

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?
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"

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

Identification and Simulation

The matching function depends on the equilibrium assignment/matching a_{ij}^E,
which depends on the whole matrix of actual z's.
Problem 1: the z's include characteristics unobserved by the econometrician, even with frictionless markets.
$\longrightarrow$ we restrict their variation and we get the equilibrium U^E and V^E functions
Problem 2: to simulate the response of the marriage market equilibrium to e.g. changes in policy, we need to work from the true primitives (the z's)
(as opposed to the equilibrium U and V).

Equilibrium conditions

Equilibrium implies that
i is matched with j iff

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

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.

The Crucial Slide

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 within each group.**

What this buys us

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.**

More Directly Useful

Necessary and sufficient conditions for men to be stable

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

$$\begin{aligned} \varepsilon_i(G,H) - \varepsilon_i(G,K) &\geq U(G,K) - U(G,H) \text{ for all } K \quad (2) \\ \varepsilon_i(G,H) - \varepsilon_i(G,0) &\geq U(G,0) - U(G,H); \quad (3) \end{aligned}$$

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

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

These are just the defining equations for a multinomial choice model.

Why not a multilogit then?

What do we need to assume about the U and V functions and the joint distribution of the ε_i and η_j?
Given covariates that vary within groups, not much (see later)
But assuming that the ε's and η's are also independent, homoskedastic, type-I extreme value is a natural starting point.

Choo and Siow (*JPE* 2006)

When there are only group covariates multilogit also 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,

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

Introducing covariates

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.)

Identification and Estimation

Within-group covariates help for identification; we mostly need

- ε_i and η_j independent for all i and j
- separate normalizations for men and women within each group:

$$\text{for all} G, \qquad U(G, 0) = 0;$$

$$\text{for all} H, \qquad V(0, H) = 0;$$

- heteroskedasticity does not depend too much on covariates.

But estimation now does not lead to closed-form formulæ.

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 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

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 ε_{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 $a_i * c_i$, $a_i^2 * c_i$ and $a_i * c_i^2$) allows for changes in the age profile of group preferences across cohorts;
- 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.

On the US

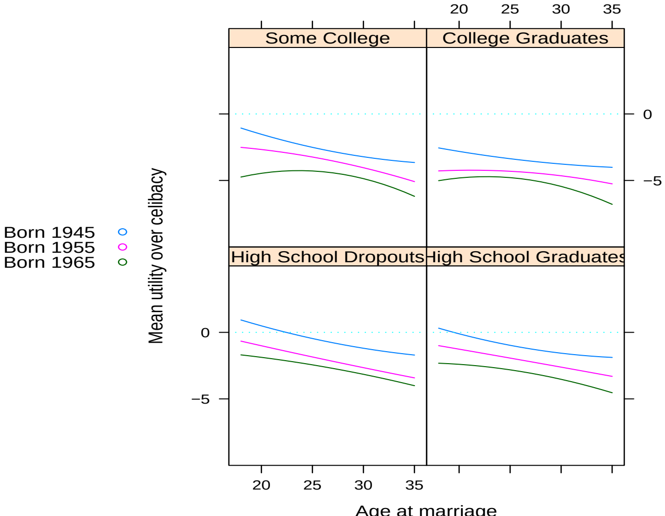
Has more preference for assortative matching amplified the increases in wage inequality?
We use the June CPS, 6 waves from 1971 to 1995.
We focus on whites (and white-white matches: 95% of matched whites have a white spouse.)
4 educational levels:

- 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).

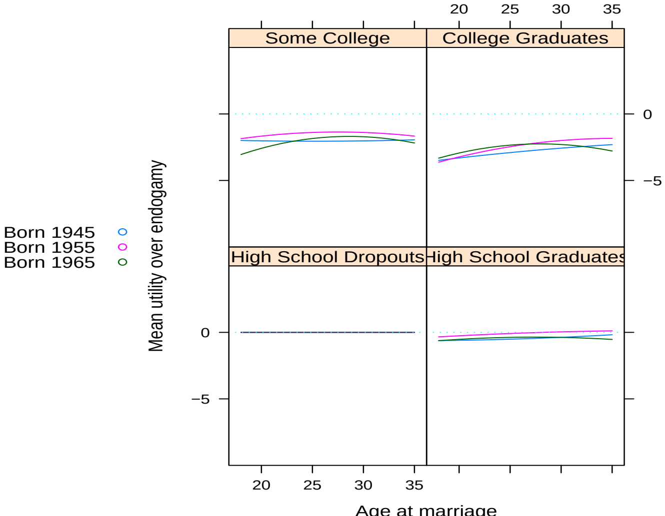
Mean Utilities

We estimate separately the models for men and women of each group $\longrightarrow$ 8 models with 44 parameters each.
Focus here on “extreme” educational levels: high school dropouts and college graduates.
First we plot the estimated $U(G, H; a, c)$ for the cohorts 1945, 1955, 1965 between ages 18 and 35.

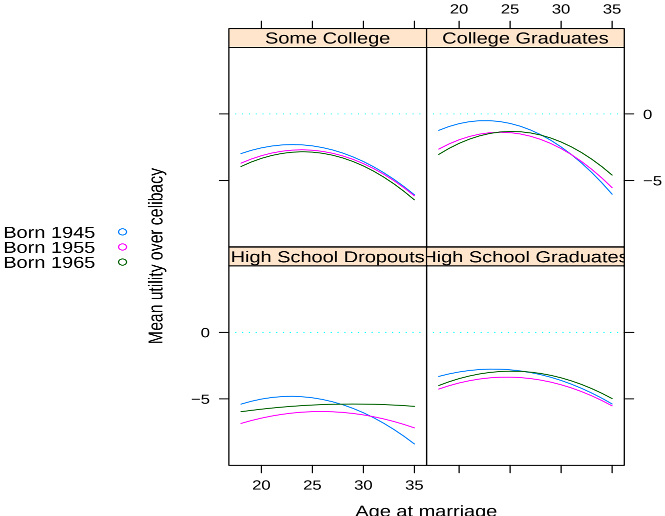
Women High School Dropouts: Mean Utility from Marrying



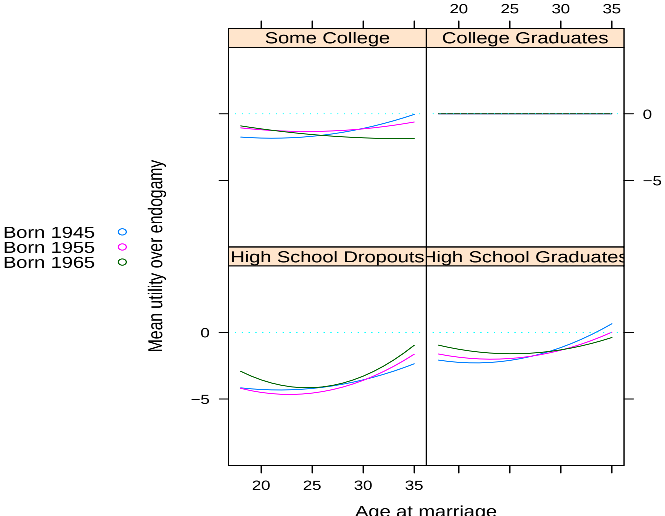
Women High School Dropouts: Mean Utility from Exogamy



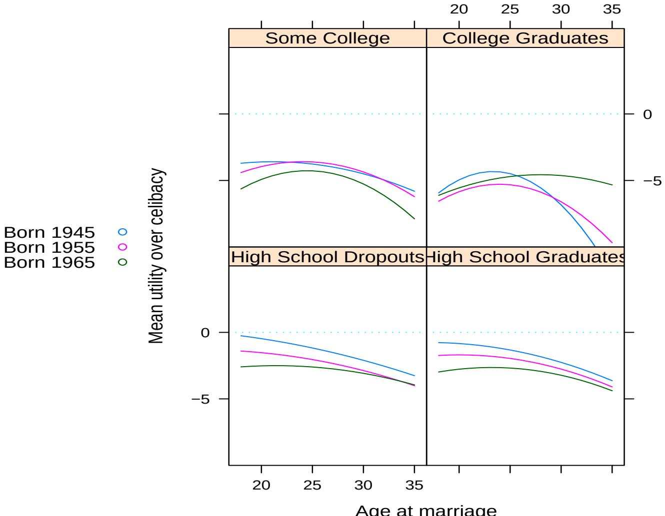
Women College Graduates: Mean Utility from Marrying



Women College Graduates: Mean Utility from Exogamy



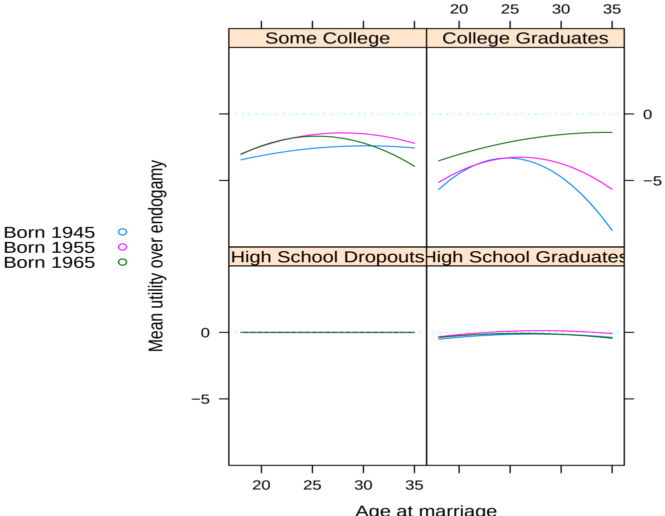
Men High School Dropouts: Mean Utility from Marrying



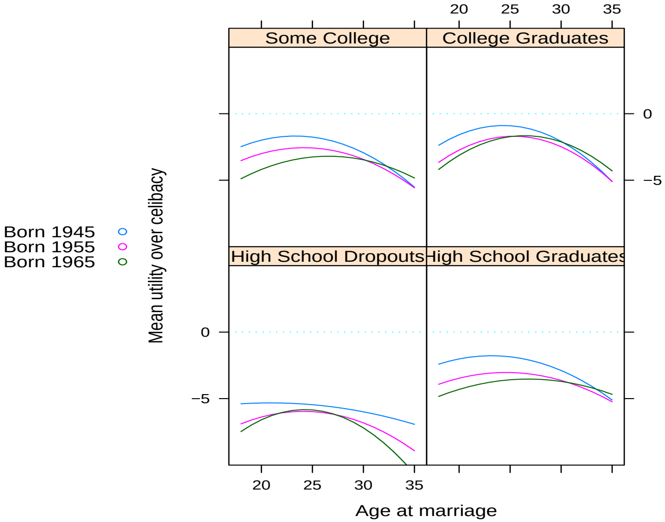
Mean Utilities from Heterogamy

Homogamy is $G = H$, heterogamy is $G \neq H$, so
we now plot the estimated $U(G, H; a, c) - U(G, G; a, c)$ for the
cohorts 1945, 1955, 1965 between ages 18 and 35.

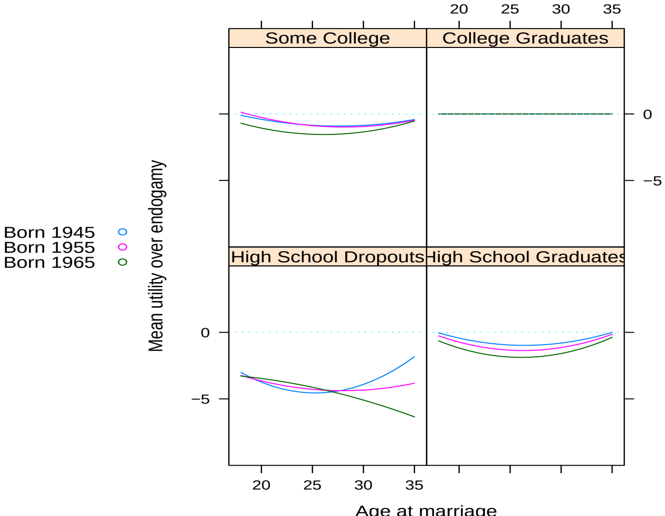
Men High School Dropouts: Mean Utility from Heterogamy



Men College Graduates: Mean Utility from Marrying



Men College Graduates: Mean Utility from Heterogamy



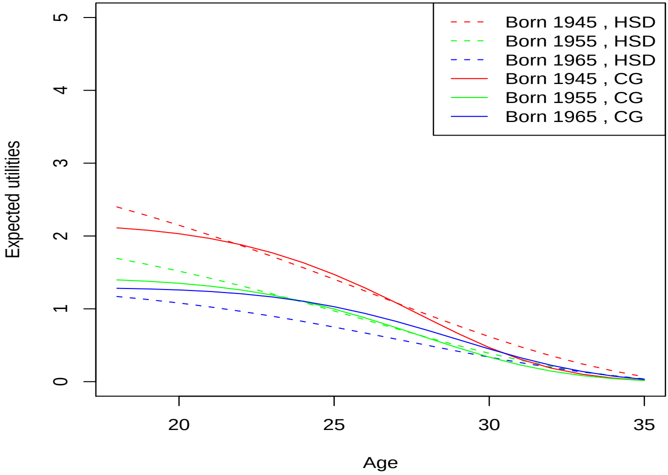
Expected returns on the Marriage Market

A man of cohort c and group G at age a can expect (not conditioning on his past!)

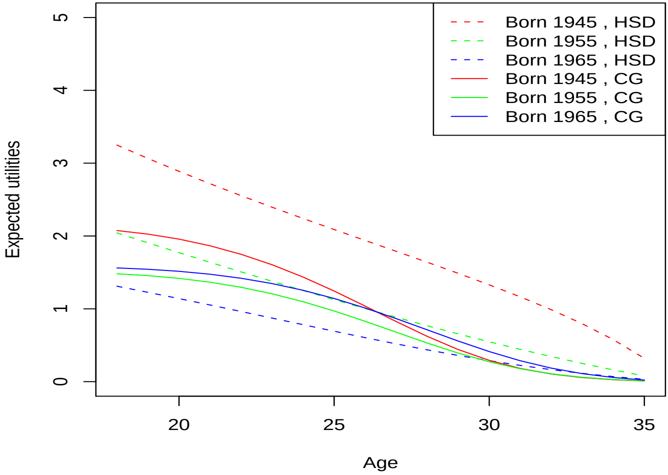
$$\overline{U}(G; a, c) = E \max_{b=a,\ldots,35; H=0,\ldots,4} (U(G, H; b, c) + \text{error})$$

This is what we plot now, for the cohorts 1945, 1955, 1965 between ages 18 and 35 and the groups $G = 1$ and $G = 4$.

Men: Expected Return on the Marriage Market



Women: Expected Return on the Marriage Market



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 ($\simeq$ 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

Source	Partial SS	df	F
Model	30265.18	615	11.30
G_i	1815.46	7	59.55
c_i	1151.17	34	7.77
G_j	1471.36	8	42.23
$G_i * G_j$	16428.43	56	293.36
$G_i * c_i$	1609.26	238	6.76
$G_j * c_i$	1948.66	272	7.16
Residual	6497.97		$R^2 = 0.8232$

ANOVA summary

Different groups have markedly different preferences, and these differences are very stable over time.

To Do

Use the model for simulations: the u's and v's depend on the actual equilibrium, but we can reconstitute the z's, and they are primitives.

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?)