Who married, (to) whom, and where?

Trends in marriage in the United States, 1850-1940*

Claudia Olivetti
Boston College and NBER

M. Daniele Paserman
Boston University and NBER

Laura Salisbury
York University and NBER

E. Anna Weber
Boston University

January 2019

Abstract

This paper presents a novel analysis about marriage in the United States in the late 19th and early 20th centuries, and its relation to socioeconomic status. We document the following facts: 1) Already in the mid-19th Century there was a socioeconomic gradient in marriage rates – men and women born to families in the bottom quartile of the occupational earnings distribution were more likely to marry than those in the top quartile. The gradient had grown steeper by the middle of the 20th Century. 2) Age at marriage follows an inverted U-shape, and exhibits a clear socioeconomic gradient, which becomes steeper over time. 3) There is a substantial increase in the degree of assortativeness by socioeconomic status over this period. 4) The mean age gap between spouses declines over time but it explains very little of the change in assortativeness by socioeconomic status. 5) Along with the socioeconomic gradient in marriage outcomes, there is also a geographic gradient: the South has more “traditional” marriage patterns than the Northeast – higher marriage rates, younger age at first marriage, more assortativeness, and a larger age gap between spouses. The geographic gradient also increases over time. The overall picture is one of a society that was becoming more segmented along the marriage dimension.

We further investigate how much of this increased segmentation can be explained by income divergence across geographic reasons. We find that regional divergence explains about one half of the socioeconomic divergence in the probability of marriage, and almost all of the increase in marital sorting. Geographic differences in urbanization rates account for much of these differences; on the other hand, geographic differences in percent manufacturing, access to railroads and scholarization rates play a smaller role.

Keywords: Marriage, Assortative Mating, Gender, Intergenerational Mobility, Regional Convergence.

JEL codes: J12, J62, N31, N32, N91, N92

*VERY PRELIMINARY AND INCOMPLETE
Who married, (to) whom, and where?  
Trends in marriage in the United States, 1850-1940

Claudia Olivetti¹  M. Daniele Paserman²  
Laura Salisbury³  E. Anna Weber⁴

¹Boston College and NBER  
²Boston University and NBER  
³York University and NBER  
⁴Boston University

January 2019
Introduction

There has been a dramatic increase in inequality over the past 40-50 years. A **New Gilded Age**?

Abundant evidence that cross-sectional inequality goes together with **lower** intergenerational mobility.
- across countries (Solon, 2002; Corak, 2006)
- across cities in the US (Chetty et al. 2014)

What is the role of the **family**?
Patterns in marriage can potentially amplify the link between inequality and mobility:
- If marriage (and in-wedlock birth) positively correlated with SES.
- If couples are positively sorted.
- Advantage conferred to children likely magnified and more persistent.

Evolution of inequality across generations depends on who marries, and who marries whom.

What do we know about mobility and the evolution of marriage over the past 150 years?
In the US, tolerance for high inequality sometimes explained by the belief that mobility is also high (Alesina et al., 2004).

In fact, mobility in the US is among the lowest in OECD (Corak, 2011).

Long and Ferrie (2013) date the “end of American exceptionalism” to the early decades of the 20th Century. They also provide evidence of multigenerational persistence in the second half of the 19th Century (Long and Ferrie, 2018).

Based on linked fathers/sons by first and last name across Census decades.

Clark (2014): evidence of low social mobility in the very long run based on the persistence of certain surnames among elite groups.

Most of this work is about the male line because daughters change last name upon marriage.
Historically, women with more education have been the least likely to marry. Today, they are more likely to be married (eventually) and less likely to divorce (Isen and Stevenson, 2010; Bailey et al. 2014).

Evidence of positive educational sorting in the US (and other countries) but some disagreement on:
- Whether it has increased in the past 50 years (Mare, 2016; Gihleb and Lang, 2016).
- Whether it has contributed to increasing income inequality (Greenwood et al., 2014; Eika et al., 2016)

Limited evidence on assortative mating prior to 1940.

Was First Gilded Age (also?) characterized by high levels of assortativeness?

- Exploit first-name based methodology (Olivetti and Paserman, 2015; Olivetti, Paserman and Salisbury, 2018) to identify socioeconomic status.

- Outcomes: Probability of marriage, age at marriage, assortative mating, spousal age gap.

Main result: divergent marriage patterns by SES.

Why divergence?

Role of geographic differences: between and within-region.

Which specific economic and demographic factors contribute the most?

- Exploit first-name based methodology (Olivetti and Paserman, 2015; Olivetti, Paserman and Salisbury, 2018) to identify socioeconomic status.

- Outcomes: Probability of marriage, age at marriage, assortative mating, spousal age gap.

Main result: **divergent** marriage patterns by SES.

- Exploit first-name based methodology (Olivetti and Paserman, 2015; Olivetti, Paserman and Salisbury, 2018) to identify socioeconomic status.

- Outcomes: Probability of marriage, age at marriage, assortative mating, spousal age gap.

Main result: **divergent** marriage patterns by SES.

Why divergence?

- Role of geographic differences: **between** and **within**-region.

- Which specific economic and demographic factors contribute the most?

- Mostly accounted for by regional differences in economic development.
- Increasing returns to human capital also played a role.
- Other mechanisms (changes in fertility, migration, investment in public schooling) less likely to matter.

- Mostly accounted for by regional differences in economic development.
- Increasing returns to human capital also played a role.
- Other mechanisms (changes in fertility, migration, investment in public schooling) less likely to matter.

Olivetti, Paserman and Salisbury (2018): stratification in marriage might amplify cross-sectional inequality and lead to lower mobility, even in a context in which married women don’t work.

- Both paternal and maternal grandparents matter for grandchildren’s outcomes, even after controlling for the father’s status.
- Transmission follows gendered lines: grandsons more strongly correlated with paternal grandfathers, maternal grandfathers more correlated with granddaughters.
How do we measure SES in pre-1940 data?
- No information about education or income.
- Occupational status? OK for men, but married women typically did not work, no occupational status.

Instead: use socioeconomic status of men’s and women’s fathers.
- Study marriage rates, sorting based on family of origin (similar to Charles et al., 2016).
- An advantage: results not sensitive to changes in marginal distribution of educational attainment.

Ideal data: individually linked data with good measures of SES and intergenerational links. Not the data we have.

How do we know SES of fathers?
Methodology

- Solution: Impute father’s economic status by **first names** (Olivetti and Paserman, 2015).
  - Take children aged 0-15 in year $t$.
  - For every first name, calculate average father’s log occupational score.
  - Use this to impute parental SES to adults with that first name in either $t + 20$ or $t + 30$.

- Key assumption: Names carry information about socio-economic status.
  - In our data: 10-17% of total variation in father’s socioeconomic status can be explained by the variation *between* names given to their children (OP 2015).
Methodology

- Nice feature: methodology is gender-neutral!

Abigail & Adam, age 30-45, 1880

$y_f$: mean log occ. earnings of fathers of daughters 0-15 named Abigail, 1850

$y_m$: mean log occ. earnings of fathers of sons 0-15 named Adam, 1850
Basic Idea

- Suppose that high SES adults call their children Adam and Abigail, low SES call their children Zachary and Zoë.

- Are Abigails more or less likely to marry than Zoës?
- Sorting: Are Abigails more likely to marry Adams (high degree of sorting), or equally likely to marry Adams or Zacharys (low degree of sorting)?
- Additional questions: age gaps between spouses, age at first marriage, etc.

- Allows marriage patterns by socioeconomic status of family of origin going far back in time.
Caveats

Potential biases of our estimates:

- **Attenuation bias**: first-name based measure is an imperfect proxy of SES.
- **Upward bias**: possible if first names have direct effect on marriage outcomes.

Interpreting trends: What if names becomes more informative over time?

- Previous work (OP2015): only limited evidence of increased informativeness of names.
- But needs further analysis.
**Data**

- IPUMS 1% samples (whites only) from 1850-1930, 1% extract from full count 1940 Census (with names).

- Different 15 year birth cohorts of men & women observed as 30-45 year olds.
  - For some analyses, also 20-35 year olds.

- Robustness (for now):
  - Alternative definition of income: Preston-Haines with farm adjustment (national, by state)
  - Real estate wealth using 1850 full count Census
  - Personal wealth using 1860 and 1870 Census
  - LIDO
  - Linked data for men
  - Assign income by name-region
  - Use full count 1940 census retrospectively
  - Soundex
Data

- IPUMS 1% samples (whites only) from 1850-1930, 1% extract from full count 1940 Census (with names).
- Different 15 year birth cohorts of men & women observed as 30-45 year olds.
  - For some analyses, also 20-35 year olds.
- Robustness (for now):
  - Alternative definition of income:
    - Preston-Haines with farm adjustment (national, by state)
    - Real estate wealth using 1850 full count Census
    - Personal wealth using 1860 and 1870 Census
    - LIDO
  - Linked data for men
  - Assign income by name-region
  - Use full count 1940 census retrospectively
  - Soundex
Marital Status by SES

Fraction Ever Married By Parental Occupational Income Quartile

Men, Ages 30-45

Women, Ages 30-45

1840 1860 1880 1900
15 year birth cohort

Fraction

.75 .8 .85 .9

Quartile 1 (lowest) Q2 Q3 Q4

Quartile 1 (lowest) Q2 Q3 Q4
Marital Status by Region of Birth

**Fraction Ever Married By Region of Birth**

**Men, Ages 30-45**

**Women, Ages 30-45**
Marital Status - Summary

- SES gradient in fraction married, for both women and men.
  - What's up with men?

- Becomes steeper for women over time, not much going on for men.

- Regional differences: divergence in marriage patterns between Northeast and South, for both men and women.

- How large is the gap? We can compare to Bailey et al. (2014).
Comparison to Bailey et al. (2014) - Fraction Married

By own education

Not bad, given different samples and definitions of SES!
Who Marries Whom?

- We have shown evidence of increasing gaps in marital status and age at marriage by both SES and region.

- Is there also divergence in who marries whom?

- Two measures of assortative mating:
  - Fraction women marrying men at different quartiles of parental SES.
  - Correlation in parental SES between spouses.
Assortative Mating - Sorting by SES quartiles

Fraction Married to Men in Bottom Quartile, Women ages 30-45

Fraction Married to Men in Top Quartile, Women ages 30-45

Sorting, far from perfect homogamy, increasing gaps especially for Q1 women
Assortative Mating - Parental SES Correlations

Correlation: Parental Log Occupational Income

Husbands & Wives, Ages 20-35

Husbands & Wives, Ages 30-45
Clear increase in correlation of parental log occupational scores.

But magnitude appears very small: 0.04-0.09

In modern data:
- Correlation in spouses’ educational attainment $\approx 0.6$ (Gihleb and Lang, 2016).
- Correlation in spouses’ parental wealth $\approx 0.4$ (Charles et al., 2016).

Based on simple model of marriage and intergenerational mobility: given imperfect assortative mating, and imperfect measure of SES, magnitudes are not unreasonable.
Within regions, correlation is lower, and trends differ.
Positive assortative mating overall, and increase over time.

Correlations small, but probably attenuated, and not unreasonable.

Within-region assortativeness smaller, and different trends across regions.
Age Gap between Spouses

- In more traditional societies, women tend to marry older men.
- Older husbands perhaps more financially stable, but lower match quality?
- What is the relationship between SES and the age gap? Has it changed over time?
Spousal Age Difference by Parental Income Quartile

Women, Ages 30-45

Age Gap

15 year birth cohort

Quartile 1 (lowest)  Q2  Q3  Q4
Age Gap by Region

Spousal Age Difference By Region (Husband-Wife)

Women, Ages 30-45

15 year birth cohort

1840 1860 1880 1900

Age Gap

Northeast  Midwest  South

26/ 44
Consistent pattern of divergence in marriage outcomes (ever married, age gap, spousal SES) by SES.

In large part, mirrored by divergence in marriage patterns across regions.
- At a time when regions were also growing apart in economic outcomes (Kim and Margo, 2004; Lindert and Williamson, 2016).

Questions:
- How much of the divergence by SES is explained by regional divergence?
- Which specific factors explain the divergence?
Econometric Framework

- Regression framework:

\[
Y_{iqst} = \sum_{q=2}^{4} \sum_{t=1880}^{1940} \gamma_{qt} (Q_{iq} \times T_{it}) + \tau_t + \delta' Z_{iqst} + u_{iqst}
\]  

(1)

\[
Y_{iqst} = \sum_{q=2}^{4} \sum_{t=1880}^{1940} \gamma_{qt} (Q_{iq} \times T_{it}) + \tau_t + \zeta_{st} + \delta' Z_{iqst} + u_{iqst}
\]  

(2)

- \(Y_{iqst}\): outcome for individual \(i\), in quartile \(q\), in state \(s\), at time \(t\)
- \(Z_{iqst}\): individual level controls (age, foreign born, urban status).
- \(\tau_t\): time fixed effect
- \(Q_{iq}\): SES quartile dummies; \(T_{it}\): time dummies
- \(\zeta_{st}\): birth state-by-time fixed effect
- \(\gamma_{qt}\): the gradient between quartile \(q\) and quartile 1 in time \(t\)
Econometric Framework

- Regression framework:

\[
Y_{iqt} = \sum_{q=2}^{4} \sum_{t=1880}^{1940} \gamma_{qt} (Q_{iq} \times T_{it}) + \tau_t + \delta' Z_{iqst} + u_{iqst}
\]  

(1)

\[
Y_{iqt} = \sum_{q=2}^{4} \sum_{t=1880}^{1940} \gamma_{qt} (Q_{iq} \times T_{it}) + \tau_t + \zeta_{st} + \delta' Z_{iqst} + u_{iqst}
\]  

(2)

- Two versions: with and without state of birth fixed effects and state × time interactions.

- Main question:
  - How do \(\gamma_{qt}\) change with inclusion of place of birth fixed effects?
This chart shows the estimated coefficients and confidence intervals for $\gamma_{qt}$ from the estimation of equation (1) on slide 29. Coefficients are analogous to the gradient between Q1 and other quartiles from slide 14.
Regression Analysis: SES Divergence in Spouse’s Parental Income

Effect of Parental SES on Spouse’s Parental Log Occscore
Women, 30-45

This chart shows the estimated coefficients and confidence intervals for $\gamma_{qt}$ from the estimation of equation (1) on slide 29.
This chart shows the estimated coefficients and confidence intervals for $\gamma_{4t}$ from the estimation of equations (1) [blue circles] and (2) [grey diamonds] on slide 29. Only the gradient between Q1 and Q4 is shown. This chart shows how this gradient changes with the addition of birthplace controls.
This chart shows the estimated coefficients and confidence intervals for $\gamma_{4t}$ from the estimation of equations (1) [blue circles] and (2) [grey diamonds] on slide 29. Only the gradient between Q1 and Q4 is shown. This chart shows how this gradient changes with the addition of birthplace controls.
Geographic factors explain:
- About one half of the change in the SES-Fraction Married gradient
- Almost all of the change in the SES-Spousal SES gradient.

Which state characteristics (if any) can explain the drop?
- We will replace state and state×time fixed effects with state-level determinants of marriage outcomes
- Which characteristics have bite?
- But first, show how these characteristics are related to marriage outcomes in panel setting.
We confirm that many of these economic and demographic characteristics at the state level do have an impact on marriage outcomes, even controlling for father’s income.
State Characteristics and Geography: Regressions

\[
Y_{iqst} = \sum_{q=2}^{4} \sum_{t=1880}^{1940} \gamma_{qt} (Q_{iq} \times T_{it}) + \tau_t + \delta' Z_{iqst} + u_{iqst} \quad (1)
\]

\[
Y_{iqst} = \sum_{q=2}^{4} \sum_{t=1880}^{1940} \gamma_{qt} (Q_{iq} \times T_{it}) + \tau_t + \zeta_{st} + \delta' Z_{iqst} + u_{iqst} \quad (2)
\]

\[
Y_{iqst} = \sum_{q=2}^{4} \sum_{t=1880}^{1940} \gamma_{qt} (Q_{iq} \times T_{it}) + \tau_t
\]
\[+ \sum_{t=1880}^{1940} \beta'_{st} (X_{st} \times T_{it}) + \delta' Z_{iqst} + u_{iqst} \quad (3)
\]

- Repeat the regression analysis, but replace state-time fixed effects $\zeta_{st}$ with a state characteristic, $X_{st}$, allowing the impact of this characteristic to vary with time.
- Again, the question is how does $\gamma_{qt}$ change when we do this replacement.
This chart shows the estimated coefficients and confidence intervals for $\gamma_{4t}$ from the estimation of equations (1) [blue circles], and (2) [grey diamonds] on slide 36. Only the gradient between Q1 and Q4 is shown. This chart shows how this gradient changes with the addition of birthplace controls.
This chart shows the estimated coefficients and confidence intervals for $\gamma_{4t}$ from the estimation of equations (1) [blue circles], (2) [grey diamonds], and (3) [red squares] on slide 36. Only the gradient between Q1 and Q4 is shown. This chart shows how this gradient changes with the addition of birthplace controls and with a control for the urbanization of the woman's birth state.
State Characteristics and Geography: Spouse’s Parental Income

This chart shows the estimated coefficients and confidence intervals for $\gamma_{4t}$ from the estimation of equations (1) [blue circles], and (2) [grey diamonds] on slide 36. Only the gradient between Q1 and Q4 is shown. This chart shows how this gradient changes with the addition of birthplace controls.
This chart shows the estimated coefficients and confidence intervals for $\gamma_{4t}$ from the estimation of equations (1) [blue circles], (2) [grey diamonds], and (3) [red squares] on slide 36. Only the gradient between Q1 and Q4 is shown. This chart shows how this gradient changes with the addition of birthplace controls and with a control for the urbanization of the woman’s birth state.
We repeat this analysis with a number of other state characteristics. Controlling for the percent foreign-born yields similar results to percent urban (as these two characteristics are correlated), but other controls appear to play a smaller role.

- **Ever Married**
  - Manufacturing
  - Railroad Access
  - Percent Foreign-Born
  - Percent Male
  - Percent in School

- **Spouse’s Parental Income**
  - Manufacturing
  - Railroad Access
  - Percent Foreign-Born
  - Percent Male
  - Percent in School
Discussion and Conclusion

- Widening gradient in marriage outcomes by parental SES status for women.
  - Increasing low SES - high SES gap in marriage probability.
  - Increasing correlation in parental log occupational scores across spouses.

- Why?
  - Geography matters: Regional divergence can explain most of the increase in sorting and part of the divergence in probability of marriage.
  - Preliminary analysis: increasing urbanization seems important, manufacturing not as much.
Discussion and Conclusion

What’s next?

- Further exploration of economic and demographic factors.
  - More work on understanding patterns for men.
  - How much are the results driven by changing naming patterns (including immigration, ethnic names, etc.)?

- Additional robustness checks (*a classic!*)

- Can we exploit plausibly exogenous changes in institutions (women’s property rights, schooling, etc.) to identify causal mechanisms?
Thank You!
Age at Marriage

- Age at marriage asked in 1910, 1930 and 1940 Census years, but slightly different samples and questions.
  - 1910: Women on first marriage.
  - 1930: Currently married women.
  - 1940: Ever married sample line women (restrict to currently married for comparability)
1910 sample includes individuals still in their first marriage, 1930 sample includes all currently married individuals, 1940 sample includes sample line currently married women.
Similar pattern of divergence in age at marriage, both by SES and by region.
Again, very similar estimate for Q1, but Q4 underestimated
Generation 0: \( y_0^M \) and \( y_0^F \) are the outcomes (income, occupational status, human capital...) of fathers of men and women in generation 1.

\[
\begin{align*}
y_0^M & \sim N(\mu_0, \sigma_0^2) \\
y_0^F & \sim N(\mu_0, \sigma_0^2)
\end{align*}
\]

Intergenerational transmission:

\[
\begin{align*}
y_1^M & = \alpha + \beta y_0^M + u_M, \quad u_M \sim N (0, \sigma_u^2) \\
y_1^F & = \alpha + \beta y_0^F + u_F, \quad u_F \sim N (0, \sigma_u^2)
\end{align*}
\]
Matching process: rank generation 1 men and women by their $y_1$, perfect assortative mating.

Result:

\[
p \lim Corr \left( y_1^M, y_1^F \right) = 1
\]
\[
p \lim Corr \left( y_0^M, y_0^F \right) = \beta^2.
\]
A Note about Magnitudes

- Even if men and women are perfectly assortatively matched on \( y_1 \), correlation in parental income can be substantially smaller than 1.

- Extensions:
  - **Imperfect** assortative mating on \( y_1 \) → lower \( \text{Corr} \left( y_0^M, y_0^F \right) \).
  - Mating both on \( y_1 \) and \( y_0 \) → higher \( \text{Corr} \left( y_0^M, y_0^F \right) \). (In the limit, \( \text{Corr} \left( y_0^M, y_0^F \right) = 1 \) if only \( y_0 \) matters).

- Bottom line: given imperfect assortative mating, and imperfect measure of SES, magnitudes are not unreasonable.
Effect of Parental SES on Probability of Ever Marrying
Women, 30-45
Birthplace Controls

State Characteristics and Geography: Ever Married

Coeff. on top SES quartile

Birth Cohort

1840  1860  1880  1900  1920

-0.06  -0.04  -0.02  0  0.02

-0.06  -0.04  -0.02  0  0.02

No Controls
Birthplace Controls
Effect of Parental SES on Probability of Ever Marrying Women, 30-45

% in Manufacturing Control

Birth Cohort

- Coeff. on top SES quartile

- Back

- No Controls

- Birthplace Controls

- % in Manufacturing Control
Effect of Parental SES on Probability of Ever Marrying Women, 30-45

Birthplace Controls

Birth Cohort

Coef. on top SES quartile

1840 1860 1880 1900 1920

No Controls

Birthplace Controls
Effect of Parental SES on Probability of Ever Marrying Women, 30-45

Railroad Access Control

Birth Cohort

Coef. on top SES quartile

1840 1860 1880 1900 1920

No Controls  Birthplace Controls  Railroad Access Control

Back
State Characteristics and Geography: Ever Married

Effect of Parental SES on Probability of Ever Marrying Women, 30-45

Birthplace Controls

- Coeff. on top SES quartile
- Birth Cohort

No Controls
Birthplace Controls
Effect of Parental SES on Probability of Ever Marrying Women, 30-45

% Foreign Control

Birth Cohort

No Controls
Birthplace Controls
% Foreign Control
State Characteristics and Geography: Ever Married

Effect of Parental SES on Probability of Ever Marrying Women, 30-45

Birthplace Controls

Coeff. on top SES quartile

Birth Cohort

1840 1860 1880 1900 1920

No Controls
Birthplace Controls
State Characteristics and Geography: Ever Married

Effect of Parental SES on Probability of Ever Marrying Women, 30-45

% Male Control

Coeff. on top SES quartile

<table>
<thead>
<tr>
<th>Birth Cohort</th>
<th>No Controls</th>
<th>Birthplace Controls</th>
<th>% Male Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1840</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1860</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1880</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
State Characteristics and Geography: Ever Married

Effect of Parental SES on Probability of Ever Marrying
Women, 30-45

Birthplace Controls

Birth Cohort

Coef. on top SES quartile

No Controls
Birthplace Controls
Effect of Parental SES on Probability of Ever Marrying
Women, 30-45

% Children in School Control

Coeff. on top SES quartile

Birth Cohort

1840 1860 1880 1900 1920

No Controls  Birthplace Controls
% Children in School Control
State Characteristics and Geography: Spouse’s Parental Income

Effect of Parental SES on Spouse's Parental Log Occscore

Women, 30-45

Birthplace Controls

Birth Cohort

No Controls

Birthplace Controls
State Characteristics and Geography: Spouse’s Parental Income

Effect of Parental SES on Spouse's Parental Log Occscore
Women, 30-45

% Children in School Control

Birth Cohort

-0.01 0 0.01 0.02 0.03

Coef. on top SES quartile

1840 1860 1880 1900 1920

Birthplace Controls
No Controls
% Children in School Control
State Characteristics and Geography: Spouse’s Parental Income

Effect of Parental SES on Spouse's Parental Log Occscore
Women, 30-45
Birthplace Controls

Birth Cohort

Coef. on top SES quartile
0
-0.01
-0.02
-0.03

1840
1860
1880
1900
1920

No Controls
Birthplace Controls
State Characteristics and Geography: Spouse’s Parental Income

Effect of Parental SES on Spouse's Parental Log Occscore
Women, 30-45

Railroad Access Control

Coef. on top SES quartile

Birth Cohort

-0.1
0
0.1
0.2
0.3

1840 1860 1880 1900 1920

No Controls
Birthplace Controls
Railroad Access Control
State Characteristics and Geography: Spouse’s Parental Income

Effect of Parental SES on Spouse's Parental Log Occscore
Women, 30-45
Birthplace Controls

No Controls
Birthplace Controls
State Characteristics and Geography: Spouse’s Parental Income

Effect of Parental SES on Spouse's Parental Log Occscore
Women, 30-45
% Foreign Control

Coef. on top SES quartile

Birth Cohort

No Controls
Birthplace Controls
% Foreign Control
State Characteristics and Geography: Spouse’s Parental Income

Effect of Parental SES on Spouse's Parental Log Occscore
Women, 30-45

Birthplace Controls

Birth Cohort

No Controls
Birthplace Controls
Effect of Parental SES on Spouse's Parental Log Occscore
Women, 30-45

% Male Control

Coefficient on top SES quartile

Birth Cohort

No Controls
Birthplace Controls
% Male Control

1840 1860 1880 1900 1920
State Characteristics and Geography: Spouse’s Parental Income

Effect of Parental SES on Spouse's Parental Log Occscore
Women, 30-45

Birthplace Controls

Birth Cohort

-0.1 0 0.1 0.2 0.3
Coeff. on top SES quartile

1840 1860 1880 1900 1920
Birth Cohort

No Controls
Birthplace Controls
State Characteristics and Geography: Spouse’s Parental Income

Effect of Parental SES on Spouse's Parental Log Occscore
Women, 30-45

% in Manufacturing Control

Birth Cohort

- Coeff. on top SES quartile

1840 1860 1880 1900 1920

No Controls
Birthplace Controls
% in Manufacturing Control
Table 2: Common Names Given to Children, Ranked by Mean Father's Occupational Income, 1850-1920.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Males</th>
<th>Least Prestigious</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Edward, Walter, Harry, Paul, Donald, Abraham, Jerome</td>
<td>Jesse, Levi, Jesse, Luther, Luther, Jessie, Willie</td>
</tr>
<tr>
<td>2</td>
<td>Frederick, Frank, Walter, Harry, Kenneth, Max, Irving</td>
<td>Hiram, Isaac, Franklin, Ira, Dewey, Otis, Loyd</td>
</tr>
<tr>
<td>3</td>
<td>Edwin, Willie, Herbert, Frederick, Harold, Nathan, Jack</td>
<td>Isaac, Benjamin, Isaac, Isaac, Perry, Luther, Luther</td>
</tr>
<tr>
<td>4</td>
<td>Charles, Louis, Theodore, Ralph, Morris, Vincent, Nathan</td>
<td>Daniel, Andrew, Hiram, Willis, Virgil, Eddie, Jessie</td>
</tr>
<tr>
<td>5</td>
<td>Franklin, Fred, Edward, Philip, Max, Edmund, Abraham</td>
<td>David, Jacob, Martin, Charley, Ira, Charley, Otis</td>
</tr>
</tbody>
</table>
Table 2: Common Names Given to Children, Ranked by Mean Father's Occupational Income, 1850-1920.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1850</td>
</tr>
<tr>
<td></td>
<td>Emma</td>
</tr>
<tr>
<td></td>
<td>Alice</td>
</tr>
<tr>
<td></td>
<td>Anna</td>
</tr>
<tr>
<td></td>
<td>Isabella</td>
</tr>
<tr>
<td></td>
<td>Josephine</td>
</tr>
</tbody>
</table>

**Most Prestigious**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Least Prestigious</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1850</td>
</tr>
<tr>
<td></td>
<td>Sally</td>
</tr>
<tr>
<td></td>
<td>Nancy</td>
</tr>
<tr>
<td></td>
<td>Lucinda</td>
</tr>
<tr>
<td></td>
<td>Martha</td>
</tr>
<tr>
<td></td>
<td>Lydia</td>
</tr>
</tbody>
</table>

Exact name, nickname or alternative spelling appears more than once (most prestigious).

Exact name, nickname or alternative spelling appears more than once (least prestigious).

**Notes**: Entries in the table represent the five children names with the highest and lowest average father occupational score, by gender and Census year. Only names that appear at least 100 times are considered for the ranking.
<table>
<thead>
<tr>
<th>Year</th>
<th>Number of children ages 0–15 (1)</th>
<th>Number of distinct names (2)</th>
<th>Mean number of observations per name (3)</th>
<th>Percent of names that are singletons (4)</th>
<th>Percent of children with unique names (5)</th>
<th>Percent of children with names linked 20 years later (6)</th>
<th>Share with top-50 name (7)</th>
<th>Share of total variation in log earnings explained by between name variation (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1850</td>
<td>35,597</td>
<td>3,524</td>
<td>10.1</td>
<td>71.9</td>
<td>7.1</td>
<td>92.6</td>
<td>0.692</td>
<td>0.134</td>
</tr>
<tr>
<td>1860</td>
<td>48,114</td>
<td>4,083</td>
<td>11.8</td>
<td>70.5</td>
<td>6.0</td>
<td>93.7</td>
<td>0.695</td>
<td>0.111</td>
</tr>
<tr>
<td>1870</td>
<td>58,039</td>
<td>4,582</td>
<td>12.7</td>
<td>69.4</td>
<td>5.5</td>
<td>—</td>
<td>0.698</td>
<td>0.105</td>
</tr>
<tr>
<td>1880</td>
<td>75,004</td>
<td>6,589</td>
<td>11.4</td>
<td>69.4</td>
<td>6.1</td>
<td>92.9</td>
<td>0.653</td>
<td>0.112</td>
</tr>
<tr>
<td>1900</td>
<td>103,817</td>
<td>9,696</td>
<td>10.7</td>
<td>71.0</td>
<td>6.6</td>
<td>92.8</td>
<td>0.564</td>
<td>0.126</td>
</tr>
<tr>
<td>1910</td>
<td>117,612</td>
<td>9,818</td>
<td>12.0</td>
<td>69.5</td>
<td>5.8</td>
<td>94.1</td>
<td>0.534</td>
<td>0.126</td>
</tr>
<tr>
<td>1920</td>
<td>139,109</td>
<td>12,272</td>
<td>11.3</td>
<td>71.4</td>
<td>6.3</td>
<td>92.5</td>
<td>0.519</td>
<td>0.136</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1850</td>
<td>34,272</td>
<td>3,442</td>
<td>10.0</td>
<td>71.9</td>
<td>7.2</td>
<td>92.4</td>
<td>0.698</td>
<td>0.136</td>
</tr>
<tr>
<td>1860</td>
<td>46,874</td>
<td>4,488</td>
<td>10.4</td>
<td>70.7</td>
<td>6.8</td>
<td>92.8</td>
<td>0.657</td>
<td>0.132</td>
</tr>
<tr>
<td>1870</td>
<td>55,739</td>
<td>5,206</td>
<td>10.7</td>
<td>71.1</td>
<td>6.6</td>
<td>—</td>
<td>0.619</td>
<td>0.136</td>
</tr>
<tr>
<td>1880</td>
<td>72,160</td>
<td>7,161</td>
<td>10.1</td>
<td>69.0</td>
<td>6.8</td>
<td>92.0</td>
<td>0.548</td>
<td>0.133</td>
</tr>
<tr>
<td>1900</td>
<td>101,516</td>
<td>10,081</td>
<td>10.1</td>
<td>70.9</td>
<td>7.0</td>
<td>92.3</td>
<td>0.474</td>
<td>0.153</td>
</tr>
<tr>
<td>1910</td>
<td>114,074</td>
<td>10,103</td>
<td>11.3</td>
<td>69.3</td>
<td>6.1</td>
<td>93.5</td>
<td>0.473</td>
<td>0.154</td>
</tr>
<tr>
<td>1920</td>
<td>134,418</td>
<td>12,895</td>
<td>10.4</td>
<td>71.1</td>
<td>6.8</td>
<td>89.9</td>
<td>0.466</td>
<td>0.166</td>
</tr>
</tbody>
</table>

Notes: Column 7 shows the share of children that have 1 of the 50 most popular names, by gender. Column 8 shows the $R^2$ from a regression of father’s log occupational income on a full set of name dummies. Unless noted otherwise, the source for this and all following tables are the 1850–1920 Integrated Public Use Micro Samples of the US decennial population censuses (Ruggles et al. 2010).
Olivetti and Paserman (2015)  
Comparison of our ‘pseudo-panel’ estimator to standard (linked) estimator:

- Typical two-generation econometric model:
  \[
  y_{ij}^S = \beta y_{ij}^F + \lambda_j + u_{ij}; \\
  y_{ij}^F = \mu_j + z_{ij}.
  \]

- Probability limit of OLS with linked data:
  \[
  p \lim \hat{\eta}_{LINKED} = \beta + \frac{\text{Cov} (\lambda_j + \kappa_j, \mu_j)}{V(\mu_j) + V(z_{ij})} + \frac{\text{Cov} (\tilde{u}_{ij}, z_{ij})}{V(\mu_j) + V(z_{ij})}.
  \]

- Probability limit of pseudo-panel estimator:
  \[
  p \lim \hat{\eta}_{PSEUDO} = \frac{V(\mu_j)}{V(\mu_j) + E(\frac{1}{N_j})V(z'_{ij})} \beta + \frac{\text{Cov} (\lambda_j + \kappa_j, \mu_j)}{V(\mu_j) + E(\frac{1}{N_j})V(z'_{ij})}.
  \]
Olivetti and Paserman (2015)

- Summarizing:
  - Attenuation bias: replacing true G1 income with imputed value.
  - Name effects: direct labor/marriage market premium/penalty associated with given first name & ‘aspirational naming’. Pushes pseudo-panel estimator up.
  - Missing term: depends on correlation in unobservables not embodied in names (genetic ability, social capital) because father’s income is from different sample.

- In practice, using linked data:
  - Our estimator is lower by about 30%.
  - Estimates do not change when we add name fixed effects.

- Additional checks: Simulations, Modern data from Norway.
Note: The figure presents point estimates and 90% confidence intervals for the father/son and father-son-in-law intergenerational elasticities. The values on the horizontal axes represent the year from which the son's (son-in-law's) sample are drawn. The elasticities are obtained from a regression of son (son-in-law) log occupational income on imputed father's (father-in-law's) log occupational income. See text for details of the imputation procedure. Occupational income is based on average earnings in the occupation.
Norway: Cohorts born 1962-1982