Childhood environment and gender gaps in adulthood

Yulia Sidi

Applied General Exam
Department of Statistics, University of Connecticut

January 14, 2019

Outline

1. Background
2. Data
3. Results
   - Gender Gaps by Parent Income
   - Geographic Variation in Gender Gaps

Background

The following topics were assessed by researchers in the past:

- Gender gaps in adulthood in terms of employment rates, wage levels, etc.
- Gender gaps in childhood in terms of test scores, high school graduation, etc.

The current paper "Childhood environment and gender gaps in adulthood" by Chetty et al. (2016) combined these two topics by asking:

> What is the role of family characteristics and childhood environment in shaping gender gaps in adulthood?

Background

The paper presented here is part of Opportunity Insights project, whose mission is

> "to develop scalable policy solutions that will empower families throughout the United States to rise out of poverty and achieve better life outcomes"

Outline

1. Background
2. Data
3. Results
   - Gender Gaps by Parent Income
   - Geographic Variation in Gender Gaps

Data

Data:

- De-identified database of tax returns in US from 1996-2012 (IRS Databank), which was originally used by Chetty et al. (2014):

Population of interest:

- Children born between 1980-82, who are US citizens as of 2013 (approximately 10M children).
- Children linked with parents when claimed as dependents.
- Children were evaluated at age 30 for adulthood outcomes.

Data

The following three datasets used in this paper are publicly available on this website:

- **Childhood and Gender Gaps: Commuting Zone Employment Rates by Gender and Parent Income Quintile and Other Covariates**
- County Employment Rates by Gender and Parent Income Quintile and Other Covariates
- **National Employment Rates, Earnings, and Other Outcomes by Parent Percentile and Gender**

Outcomes of interest

Children:

- Employment: based on presence of W-2 form.
- Earnings: sum of all the earnings reported on individual's W-2 form.
- College attendance: based on at least one 1098-T form filed by a college, when a child was 18-23 years old.

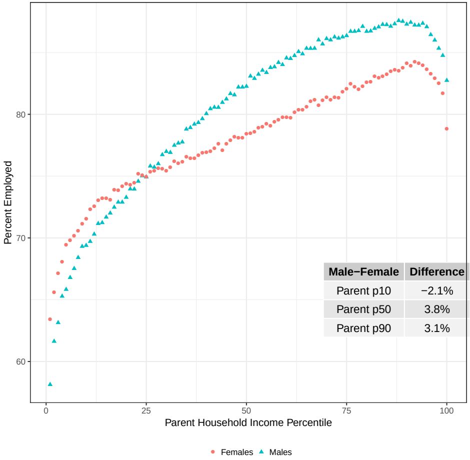
Parents:

- Household income: mean pre-tax fimily income averaged over the 5 years from 1996-2000 (1.2% of children, whose parents have zero mean income in this time period were excluded).

Outline

1. Background
2. Data
3. Results
   - Gender Gaps by Parent Income
   - Geographic Variation in Gender Gaps

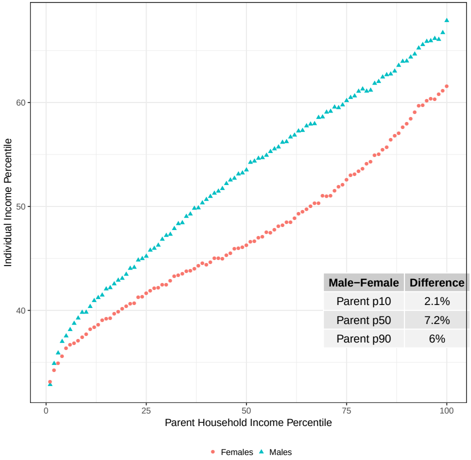
Overall Gender Gaps in Employment - Replicated



Children's Employment Rates by Parent Income Percentile

- For low income families, boys have lower employment rates than girls, while for moderate/high income families boys have higher employment rates than girls.
- This result is driven by children from single parent family (appear in the original paper).

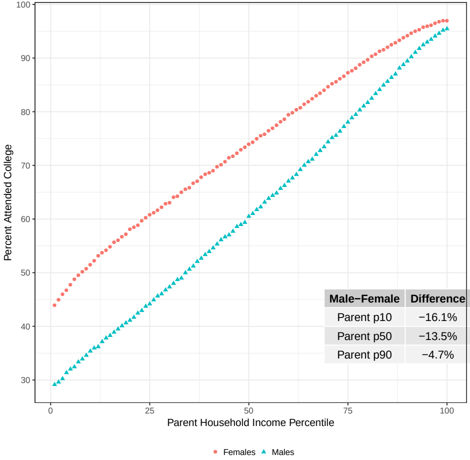
Overall Gender Gaps in Income - Replicated



Children's Income Percentile Rank (at age 30) by Parent Income Percentile

- Income percentile rank is defined as sum of wage earnings, unemployment income, disability income, and self-employment income.
- Gender gap in individual income percentile ranks is smaller for low income parents.

Overall Gender Gaps in College Attendance - Replicated



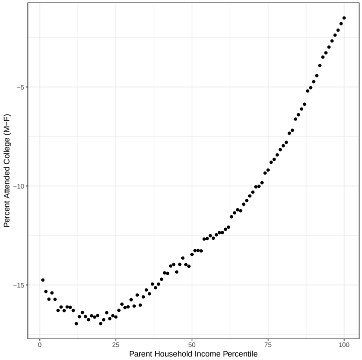
Male−Female	Difference
Parent p10	−16.1%
Parent p50	−13.5%
Parent p90	−4.7%

Children's College Attendance Rates by Parent Income Percentile

- Across the distribution of the parent's income, girls attend college at higher rates than boys.
- This gender gap is more pronounced at the lower income parents and almost disappers for the high income parents.

Overall Gender Gaps in College Attendance - Model

Gap in college attendance rates as a function of parent household income percentile (PHIP).



Gap in College Attendance Rates by Parent Income Percentile

Model 1

$$Y_i = \alpha + \beta_1 PHIP_i + \epsilon_i;\ i = 1, \ldots 100$$

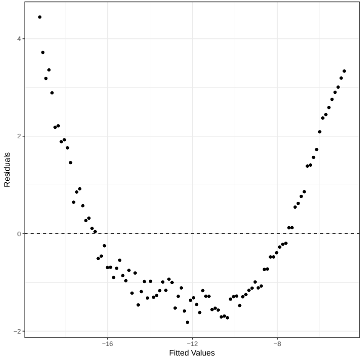
$$\epsilon_i \ iid \ N(0, \sigma^2)$$

Model 2

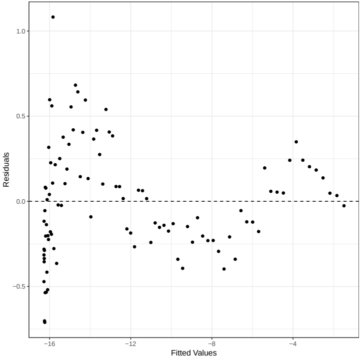
$$Y_i = \alpha + \beta_1 PHIP_i + \beta_1 PHIP_i^2 + \epsilon_i$$

$$\epsilon_i \ iid \ N(0, \sigma^2)$$

Overall Gender Gaps in College Attendance - Model Selection and Fit



Model 1: Residuals vs Fitted



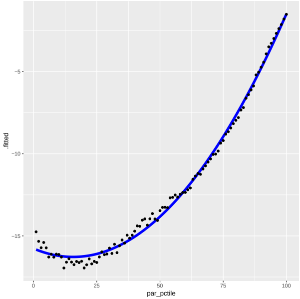
Model 2: Residuals vs Fitted

F-test resulted in $pvalue < 0.001$, suggesting we have enough evidence to believe that the squared term of PHIP needs to be included in the model.

Overall Gender Gaps in College Attendance - Model Results

Coefficient	Estimate	Std	Statistic	Pvalue
$Intercept$	-15.77	0.10	-158.70	$< .001$
$PHIP$	-0.06511	0.00454	-14.34	$< .001$
$PHIP^2$	0.0021	0.00004	47.73	$< .001$

$R^2 = 0.9949$



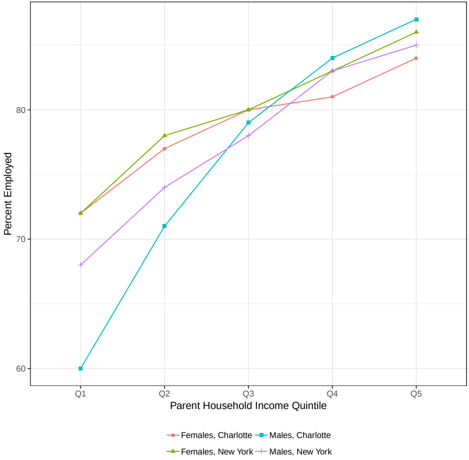
Model 2: Fitted Model

Geographic Variation in Gender Gaps

We saw variations in gender gaps as a function of parents income.
The next question is:
how these gaps vary based on where children grow up?

The rest of the analysis will be performed at the commuting zone (CZ) level. According to Chetty et al. (2016): CZs are aggregations of counties based on commuting patterns that provide a natural definition of local labor markets.

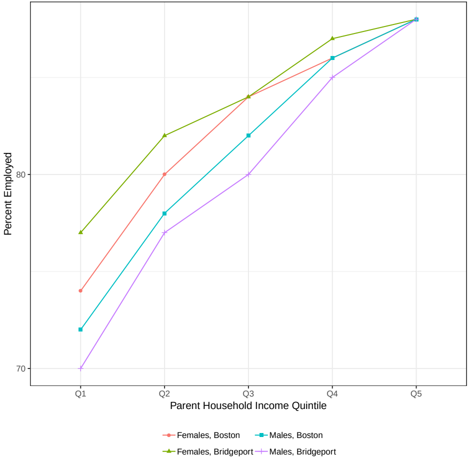
Illustrative Example of Geographic Variation - Replicated



Children's Employment Rates by Gender and Parent Income Quintile: New York vs. Charlotte CZs

- Among females, percent employment is similar between the CZs across the distribution of parent household income.
- The gaps between females and males are more pronounced in Charlotte for low-income parents.

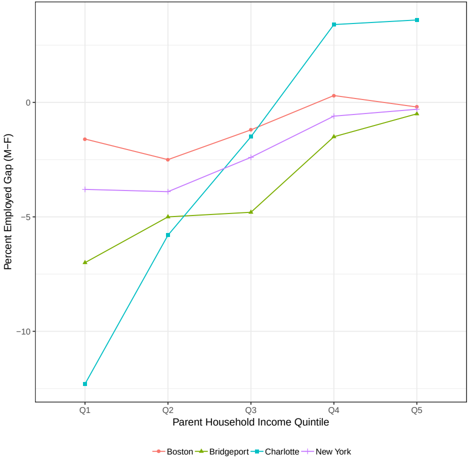
Illustrative Example of Geographic Variation



Children's Employment Rates by Gender and Parent Income Quintile: Boston vs. Bridgeport CZs

- Among females, percent employment is slightly higher in Bridgeport across the parent income distribution.
- The gaps between females and males are much more pronounced in Bridgeport for low/middle-income parents.

Illustrative Example of Geographic Variation



Gender Gaps in Children's Employment Rates by Parent Income Quintile: New-York/Charlotte/Boston/Bridgeport CZs

- The gender gaps change more extremely in Charlotte.
- In Boston, New-York and Bridgeport the gaps disappear with higher parent income, which is not the case for Charlotte.

Geographic Variation in Gender Gaps

Mean* Individual Income Ranks by Gender and Region

Region	Males	Females	Gap
Midwest	46.31	42.37	3.93
Northeast	48.24	47.02	1.22
South	43.63	41.11	2.52
West	48.28	44.23	4.05

Mean individual income ranks are defined at age 26 for children with parents at the 25th percentile of the national income distribution.

The gap in mean individual income ranks vary by region.

* Means presented here are weighted by the population size at year 2000.

Geographic Variation in Gender Gaps

The illustrative examples shown above along with individual income ranks by regions suggest that gender gaps in low-income families vary geographically.

Next we will explore what characteristics of the areas where children grow up shape the gender gaps seen in the adulthood for children from low-income families.

Missing values for the mean individual income ranks were 7% and 7.6% for males and females respectively.

Gender Gaps in Mean Income Rank- Replicated

Male-Female Difference in Mean Income Rank

Covariate	Estimate	Std	95% CI Lower	95% CI Upper
Segregation of Poverty	-2.231	0.186	-2.604	-1.858
% Black	-1.820	0.449	-2.722	-0.918
% Single Mothers	0.288	0.736	-0.498	1.075

$$Y_{ij} = \alpha + \beta_{seg(i)} Seg_{j(i)} + \beta_{blk(i)} Blk_{j(i)} + \beta_{sgl(i)} Slg_{j(i)} + State_i + \epsilon_{(i)j}$$

where $i = 1, \ldots, 51$ represents one of the states (including District of Columbia). j varied between 1-64.
CZs are clustered/nested withing a specific state $\Rightarrow$ the nested model. Thus $\epsilon_{(i)j}$ are assumed to have Normal distribution with covariance structure that takes into account the nested design together with the population weights per CZ. State effects were set as fixed effects.

Gender Gaps in Mean Income Rank- MICE

Male-Female Difference in Mean Income Rank-MICE

Covariate	Estimate	Std	95% CI Lower	95% CI Upper
Segregation of Poverty	-2.23	0.18	-2.59	-1.87
% Black	-1.816	0.446	-2.69	-1.867
% Single Mothers	0.28	0.389	-0.481	1.043

This analysis was done using predictive mean matching for clustered design with weights (Vink et al., 2015). 10 imputed datasets were used for this multiple imputation.

References

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics 129*(4), 1553–1623.

Chetty, R., N. Hendren, F. Lin, J. Majerovitz, and B. Scuderi (2016). Childhood environment and gender gaps in adulthood. *American Economic Review 106*(5), 282–88.

Vink, G., G. Lazendic, and S. van Buuren (2015). Partioned predictive mean matching as a large data multilevel imputation technique. *Psychological Test and Assessment Modeling 57*(4), 577–594.

Thank you!

Missing values in covariates

Covariate	% Missing
College tuition	21.70
College Graduation Rate	21.60
Number of Colleges per Capita	21.2
High School Dropout Rate	20
Test Score Percentile	4.9
Teenage (14-16) Labor Force Participation	4.3
Top 1% Income Share	4.3
Student Teacher Ratio	4
Violent Crime Rate	3.6
Growth in Chinese Imports	2.6
Social Capital Index	2.6
Migration Inflow Rate	2.3
Migration Outflow Rate	2.3
Local Tax Rate	0.1
Segregation of Poverty	0
Fraction Black	0
Fraction of Children with Single Mothers	0
Fraction Foreign Born	0
Fraction of Adults Divorced	0
Share Working in Manufacturing	0
Fraction of Adults Married	0
Racial Segregation	0
State EITC Exposure	0
Fraction with Commute $<$ 15 Mins	0
Gini Coefficient	0
Household Income per Capita	0
Fraction Religious	0
Tax Progressivity	0