Understanding Chetty, et al. (2014)
What Can We Conclude from an Analysis of 10 Million Intergenerational USA Tax Records?

Bradley J. Setzler

Presented to Economics 350
Department of Economics
University of Chicago
setzler@uchicago.edu

February 11, 2014
Introduction

Perhaps the most discussed topics in the economics profession, as well as in public discourse, are inequality and earnings mobility:

- I document around 60 estimates of the intergenerational elasticity of earnings (IGE) for the USA alone in the economics literature, with most of the research taking place within the past decade.
- In November, Pope Francis declared opposition to inequality an official stand of the Catholic Church, naming inequality as a source of violence.
- In December, President Obama declared inequality and social mobility "the defining challenge of our time".

Most of what is known about mobility is based on the IGE, but what is the IGE?
The IGE was originally derived by Becker and Tomes (1979) as the reduced form of a structural model of parental investments in children. However, the strong functional form assumptions describing the development of children in response to investments are not referenced or tested in the modern IGE literature.

In its present form, the IGE is the OLS estimate of the regression coefficient, $\beta$, in the equation,

$$\log Y_i^c = \alpha + \beta \log Y_i^p + \gamma X_i + \epsilon_i,$$

(1)

where $Y$ is income, $c$ denotes child, $p$ denotes parent, and $X$ is a vector of other characteristics, usually including a quadratic in the ages of parents and children.

As we saw in Becker’s (2014) presentation to this class, he remains interested in the economic origins of the IGE rather than its purely descriptive content.
I survey the distribution of USA IGE estimates in the literature, around two-thirds of which I have extracted from the survey by Corak (2006):

![Histogram of USA IGE Estimates]

- Median: 0.40
- SD: 0.13

Bradley J. Setzler, University of Chicago

Understanding Chetty, et al. (2014)
Here are USA IGE estimates since 2004 (few of which were available for the survey by Corak (2006)), when the topic found renewed popularity:
As can be seen from my survey, the median USA IGE estimate has remained 0.40, but values as low as 0.20 and as high as 0.60 remain within two standard deviations. Researchers have attributed the lack of agreement in the literature to the following econometric and data availability challenges pointed out by, e.g., Solon (1992):

- Earnings data are only available for small numbers of years, and the estimates ought to depend on both:
  - the ages of the parents and children at the time earnings are measured;
  - the number of years of earnings averaged to form lifetime earnings estimates;
- Estimates may depend on whether one or both parents are used to construct parental earnings; and,
- Estimates may depend on whether children are male, female, have siblings, are middle children, or other family structure characteristics.
Chetty, et al (2014) can be seen as an attempt to solve the data availability challenges. This work falls within the literature on descriptive IGE, as it does not attempt to attribute causality to any particular characteristic of parents or children.

Summarizing broadly, this paper repeats the analytic strategies of others with the following extensions and modifications:

- Uses administrative data from federal tax records;
- Contains a massive sample of 44 million families;
- Uses a smaller, matched sample with greater data availability to perform out-of-sample sensitivity tests; and,
- Slightly extends the empirical specification of the intergenerational earnings descriptive parameter.
Furthermore, this paper extends the idea of Corak (2013) to relate earnings mobility to inequality across geographical regions. Corak uses the survey of IGE estimates from his earlier (2006) paper to construct what Krueger (2012) named the Great Gatsby Curve:
However, Setzler (2014) finds that the Great Gatsby Curve is not robust to the choice of IGE estimates chosen from the menu of available estimates. Here is the distribution of Great Gatsby Curve slope estimates, some of which are negative:
Setzler finds that, using before tax-and-transfer Gini coefficients, before tax-and-transfer earnings, and accounting explicitly for taxes and transfers (rather than subsuming taxes and transfers into the Gini coefficient), 37.7% of possible Great Gatsby Curves have positive and statistically significant slope estimates, while 70.5% and 95.7% of taxes and transfers slope estimates are negative and statistically significant, respectively.

Furthermore, the sensitivity analysis reveals that Corak has chosen IGE’s that correspond to a Great Gatsby Curve slope estimate that is in the largest 15% using before tax-and-transfer Gini coefficients and in the largest 10% using after tax-and-transfer Gini coefficients.

The study of intergenerational inequality correlates across geographical regions would benefit from much richer data on both individual and community characteristics.
Introduction

Chetty, et al. acquire many data sets in order to explain geographic differences in intergeneration mobility by accounting for the following correlates:

- Taxes rates and transfers (public goods);
- College attainment and tuition costs;
- School district test scores and resources;
- Teenage motherhood (for females) and family structure;
- Racial composition and spacial segregation;
- Income spacial segregation and inequality;
- Occupational composition and pressure from Chinese imports;
- Migration in and out of the region; and,
- Community religiosity and crime rates (social capital).
Before proceeding to the content of the paper, a few warnings:

• This is one of the largest empirical projects ever conducted by micro-economists, with a large research staff managing over a dozen matched data sets and many millions of observations identified over time and by location;

• The sheer size and complexity of this paper can make it difficult to see the trees for the forest, so we must be diligent to step back and look for methodological errors and omissions;

• Most of the complexity comes from the data, as a dozen data sets are used for the main estimates with others added for sensitivity analysis; and,

• We will find that there is almost no theory in this paper, these are ad hoc statistical models only, and that should concern us, especially when looking for economic content.
The major innovation of this paper is data. Here, I document the data sources used, what information is taken from which data, rules used to make choices of data exclusion and coding, and possible sources of concern.

We start by discussing the primary data. The primary data are the federal tax returns of every American who filed a tax return between the years 1996 and 2012. This data includes 1040 forms, which include:

- Wages, salaries, tips, and other payments for labor;
- Dividends, capital gains, and other income from investments;
- Pensions, unemployment benefits, disability insurance;
- Marital status, dependents claimed, and head of household;
- Other taxes already paid, such as those withheld by employers;
- Various types of tax credits, including earned income; and,
- Location (ZIP code).
What matters for the study of intergenerational inequality is that the earnings of parents can be matched to the earnings of their children when the children are old enough to have earnings. Since adult children are not mentioned on parents’ tax forms, this means that records must be available both when the child is young enough to be included on the parents’ forms and when the child is old enough to have earnings.

Because the data covers a 16-year time span, and children are not reliably claimed as dependents after age 16, the only cohorts who can be matched to parents with high rates of parent matching and on whom we observe earnings after age 30 are those children who are 14-16 years of age in 1996. This restriction still permits a "core" population of 10 million.
Individual-level Data

Because the poor receive transfers from filing taxes, the authors note, almost all parents file taxes at some point. Among the core sample of children, over 95% are matched to their parents.

However, there are multiple ways to define both parents’ and children’s earnings. Their primary approach is to compute average earnings per parent over a 5 year period, permitting the number of parents to vary by year depending on dependent claims. Children’s earnings are average over the latest two years observed for the child’s household. We will see sensitivity analyses of these definitions later.
Individual-level Data

In the event of missing earnings (6% of children’s earnings are missing), the authors use related data to construct earnings. In addition to the 1040 tax data, they also have:

- Wages reported on the W-2 form;
- Unemployment benefits reported on the 1099-G form; and,
- Gross social security and disability benefits on SSA-1099 form.

Adding these together provides a proxy for income. In the event that neither 1040 nor these other forms are available they face a common problem in the IGE literature: determining how to code missing earnings data. These authors check for differences when those reporting no earnings are coded to $0, $1, or $1,000. For reasons unknown, I have found no IGE paper that uses structural wage imputation as in Heckman (1974). Lastly, note that income is top-coded at $100 million.
The last information taken from the main federal tax data are:

- College attendance, defined as the existence of a 1098-T form when the child is between 18 and 21 years of age; and,
- A woman is defined as a teenage mother if she claims a dependent while between the ages of 13 and 19.

Some characteristics of the resulting population:

- Median family income of parents is $60,129;
- Median family income of children is $34,000; and,
- Among the 30% of children matched to a single parent, 72% have female parent.
For robustness checks, they also use the Statistics of Income Sample, a cross-sectional survey conducted by the IRS prior to 1996. It includes taxpayer identification numbers for dependents of older cohorts and can thus be used to match parents in the 1040 data to their older children. A random sample of approximately 63,000 children born between 1971 and 1979 is then available for comparison with the core sample. They also use 2011-12 CPS and 2011-12 ACS for income distribution comparisons.

In summary, the authors have population-wide data on earnings, transfer benefits, and family structure of parents and their children, with at least 2 years of this information for children and many years for parents.
Individual-level Data

I have two big concerns with the use of this data:

• Their resulting measure of earnings is quite unusual. It is earnings before taxes, as in the rest of the IGE literature, but after *some* transfers (e.g., Social Security) while excluding others (e.g., TANF), unlike the other IGE literature. They perform no sensitivity analysis on the inclusion of these transfers.

• They exclude families in which the children are not citizens in the most recent data. This means they have no data on the estimated 13 million illegal immigrants in the USA (PEW 2013), nor the legal non-citizens. This will cause problems later in interpreting intergenerational inequality within geographical regions, as regions vary widely in illegal immigrant population share. They do not comment on how this exclusion may impact results.
A commuting zone (CZ) is an aggregation of counties based on interconnected economies, formed from commuting patterns constructed from the 1990 Census by Tolbert and Sizer (1996), with some updates by the authors. There are 741 CZ’s, each of which contains on average four counties and a population of 380,000 people.

Children are assigned to ZIP codes based on residence at age 16, even if they move as adults. In this sense, any information about CZ-level mobility corresponds to location where the child was raised, not where the child moved as an adult, though they have some analysis about where the children migrated as well.

Here, I document the sources of variables used in this paper:
## Commuting Zone-level Data

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2000 Census</strong></td>
<td>Number of individuals in each CZ; CZ fraction black, white, and Hispanic; Gini coefficients, fraction middle class, and top 1% income shares; CZ segregation of affluence and poverty using income distribution of households in census tracts; CZ average commuting time; CZ labor force participation; CZ fraction of workers in manufacturing; CZ fraction not born in CZ</td>
</tr>
<tr>
<td><strong>1992 Census of Government</strong></td>
<td>CZ tax rates as average tax revenue, CZ government expenditures as average spending; CZ fraction married, divorced, and single mothers among households with own children</td>
</tr>
<tr>
<td><strong>Tax Foundation</strong></td>
<td>2008 state income tax rates for incomes above $100,000 and incomes in bottom tax bracket</td>
</tr>
</tbody>
</table>
## Commuting Zone-level Data

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hotz and Scholz (2003)</td>
<td>EITC rates by state</td>
</tr>
<tr>
<td>ACCRA</td>
<td>CZ price index based on cost of living</td>
</tr>
<tr>
<td>Autor, et al. (2013)</td>
<td>CZ percentage change in Chinese imports per worker</td>
</tr>
<tr>
<td>Rupasingha and Goetz (2008)</td>
<td>SZ social capital, which is an index of participation in civic, religious, sports, political, labor union, and other institutions, from Census</td>
</tr>
<tr>
<td>Association of Religion Data Archives</td>
<td>CZ rate of religious adherents</td>
</tr>
<tr>
<td>FBI Uniform Crime Report</td>
<td>CZ rate of arrests for serious, violent crimes</td>
</tr>
</tbody>
</table>
### Commuting Zone-level Data

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCES Common Core of Data</td>
<td>CZ school expenditures per student in 1996-97; CZ student-teacher ratios for 1996-97; CZ high school dropout rates for 2000-01</td>
</tr>
<tr>
<td>IPEDS</td>
<td>CZ number of colleges per capita in IPEDS 2000; average in-state tuition and fees for first-time, full-time undergraduates at Title IV institutions; average graduate rate within 150% of normal time in IPEDS 2009</td>
</tr>
<tr>
<td>Global Report Card</td>
<td>CZ scores on National Math Percentile and National Reading Percentile for each school district relative to its state in grades 3-8 from the NAEP Exam</td>
</tr>
</tbody>
</table>
The authors adapt the rank-rank regression of Dahl and DeLeire (2008). They rank children and parents based on income, then regress child ranks on parent ranks. This differs from the usual IGE regression only in the transformation of incomes. The common specification is,

\[
f(Y_{ci}) = \alpha + \rho f(Y_{pi}) + \eta_i
\]

where \( \rho \) is estimated by OLS, and the functional form \( f \) is usually identity or log. If \( f \) is log, then \( \rho \) is interpreted as the IGE. In this paper, \( f \) is the rank in the distribution so that \( f(Y) \) has the uniform distribution. They choose the rank-rank specification after finding that the log-log specification does not fit the data well and implies IGE estimates sensitive to outliers and to choice of coding missing income (plot on next slide). They find that the rank-rank transformation fits the data well across the CZ’s.
Descriptive Parameters and Estimates

FIGURE I: Association between Children's and Parents' Income

A. Level of Child Family Income vs. Parent Family Income

B. Log Child Family Income vs. Log Parent Family Income

Notes: These figures present non-parametric binned scatter plots of the relationship between child income and parent income. Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Child income is the mean of 2011-2012 family income (when the child was around 30), while parent income is mean family income from 1996-2000. Incomes are in 2012 dollars. To construct Panel A, we bin parent family income into 100 equal-sized (centile) bins and plot the mean level of child income vs. mean level of parent income within each bin. For scaling purposes, we do not show the point for the top 1% in Panel A, as mean parent income in the top 1% is $1.4 million. In Panel B, we again bin parent family income into 100 bins and plot mean log income for children (left y-axis) and the fraction of children with zero family income (right y-axis) vs. mean parents' log income. Children with zero family income are excluded from the log income series. In both panels, the 10th and 90th percentile of parents' income are depicted in dashed vertical lines. The coefficient estimates and standard errors (in parentheses) reported on the figures are obtained from OLS regressions on the micro data. In Panel A, we report separate slopes for parents below the 90th percentile and parents between the 90th and 99th percentile. In panel B, we report slopes of the log-log regression (i.e., the intergenerational elasticity of income or IGE) in the full sample and for parents between the 10th and 90th percentiles.

Bradley J. Setzler, University of Chicago
Understanding Chetty, et al. (2014)
They consider $\rho$ a relative measure of mobility. A one point increase in the rank of parents’ income is statistically related by a $\rho$ point increase in the rank of child’s income. By contrast, they define,

$$E\left[ r \left( Y_i^c \right) \Big| r \left( Y_i^p \right) = P \right] = \alpha + \rho P, \quad (3)$$

as an absolute measure of mobility, in particular, calling “absolute upward mobility” the case where $P < 0.50$. This is the expected income rank of children whose parents are below the median income. From the uniform distribution of ranks and linearity,

$$E\left[ r \left( Y_i^c \right) \Big| r \left( Y_i^p \right) < 0.50 \right] = \alpha + 0.25\rho. \quad (4)$$
I have some concerns with these parameters and their estimation:

- It’s unclear why we would interpret a change in distribution rank as a measure of absolute mobility when conditioning on parents being below the median, but relative mobility when allowing any parental income rank. Clearly, both represent mobility relative to the within-location distributions of income.

-Validity of the estimator requires that $E[\eta_i \mid r(Y_i^p)] = 0$, and it’s unclear why this would ever be true, as we can imagine many omitted variables that are causally related to both child and parent incomes. The stronger homoskedasticity assumption on $\eta$ is needed for OLS to minimize variance. Statistical assumptions are ignored in the paper.
And this model permits relatively uninteresting inference:

- These are purely descriptive parameters - no causal model is proposed to explain why a counterfactual shift in parents’ income rank would change the income rank of children, all else equal.

  - This would imply that parental altruism/paternalism does not depend on parental income, but rather depends only on parental income relative to others. Said differently, if the rank-rank regression were the true causal model, then doubling the incomes of all parents could, for example, reduce the incomes of children, as long as child ranks remain unchanged.

- Furthermore, we cannot interpret the better fit of the rank-rank regression as proof that the constant, linear elasticity (log-log) regression is false, as the result could be driven by omitted variables that are more correlated with income levels than ranks.
When using the full data, not weighted by CZ, they find:

- The IGE is 0.344, but it rises to 0.452 when omitting the top and bottom 10%, while it rises as high as 0.618 when missing data is coded to $1;
- The coefficient of the rank-rank regression is 0.341 using 5 years of parental income and 2 years of child income;
- Using only the parent with higher income, the regression coefficient falls to 0.312;
- For male children it is 0.336 compared to 0.346 for females;
- The estimate falls to 0.282 using individual child earnings instead of child’s household, which the authors interpret as evidence of assortative mating of children;
- Figure on the next slide shows how the estimate changes as a function of child’s age - it is increasing but appears to converge at age 32 and the same for the SOI random sample.
Descriptive Parameters and Estimates

A. Lifecycle Bias: Rank-Rank Slopes by Age of Child

Notes: This figure evaluates the robustness of the rank-rank slope estimated in Figure IIa to changes in the age at which child income is measured (Panel A) and the number of years used to measure parents’ income (Panel B). In both panels, child income is defined as mean family income in 2011-2012. In Panel A, parent income is defined as mean family income from 1996-2000. Each point in Panel A shows the slope coefficient from a separate OLS regression of child income rank on parent income rank, varying the child’s birth cohort and hence the age at which child income is measured in 2011-12. The blue dots use the extended sample in the population data, while the red triangles use the 0.1% Statistics of Income stratified random sample. The first point in Panel A corresponds to the children in the 1990 birth cohort, who are 21-22 when their incomes are measured in 2011-12 (denoted by age 22 on the figure). The last point for which we have population-wide estimates corresponds to the 1980 cohort, who are 31-32 (denoted by 32) when their incomes are measured. The last point in the SOI sample corresponds to the 1972 cohort, who are 39-40 (denoted by 40) when their incomes are measured. The dashed red line is a lowess curve fit through the SOI 0.1% sample rank-rank slope estimates. In Panel B, we focus on children in the core sample (1980-82 birth cohorts) in the population data. Each point in this figure shows the coefficient from the same rank-rank regression as in Figure IIa, varying the number of years used to compute mean parent income. The first point uses parent income data for 1996 only to define parent ranks. The second point uses mean parent income from 1996-1997. The last point uses mean parent income from 1996-2012, a 17 year average.

Bradley J. Setzler, University of Chicago
Understanding Chetty, et al. (2014)
Descriptive Parameters and Estimates

Here are the resulting estimates when performing the rank-rank regression on each CZ:

B. Relative Mobility: Rank-Rank Slopes \( (\bar{y}_{100} - \bar{y}_0)/100 \) by CZ
Descriptive Parameters and Estimates

Here are the resulting estimates when performing the absolute upward mobility estimate on each CZ:

A. Absolute Upward Mobility: Average Child Rank for Below-Median Parents ($\bar{y}_{25}$) by CZ

B. Relative Mobility: Rank-Rank Slopes ($\bar{y}_{100} / \bar{y}_0$) by CZ

Notes: These figures present heat maps of our two baseline measures of intergenerational mobility by commuting zone (CZ). Both figures are based on the core sample (1980-82 birth cohorts) and baseline family income definitions for parents and children. Children are assigned to commuting zones based on the location of their parents (when the child was claimed a dependent), irrespective of where they live as adults. In each CZ, we regress child income rank on a constant and parent income rank. Using the regression estimates, we define Absolute Upward Mobility ($\bar{y}_{25}$) as the intercept + 25 ◊ (rank-rank slope), which corresponds to the predicted child rank given parent income at the 25th percentile (see Figure V). We define relative mobility as the rank-rank slope; the difference between the outcomes of the child from the richest and poorest family is 100 times this coefficient ($\bar{y}_{100} / \bar{y}_0$). The maps are constructed by grouping CZs into ten deciles and shading the areas so that lighter colors correspond to higher absolute mobility (Panel A) and lower rank-rank slopes (Panel B). Areas with fewer than 250 children in the core sample, for which we have inadequate data to estimate mobility, are shaded with the cross-hatch pattern. In Panel B, we report the unweighted and population-weighted correlation coefficients between relative mobility and absolute mobility across CZs. The CZ-level statistics underlying these figures are reported in Online Data Table V.
The correlation between upward absolute mobilities and rank-rank regressions is 0.90.

The remaining results use the CZ-level data to search for correlations among local conditions and intergenerational earnings mobility:

1. High African American fraction in CZ is negatively correlated with upward mobility, for both blacks and whites;
2. More racial segregation is negatively correlated with upward mobility, as is segregation of low-income families;
3. Commute times are strongly correlated with upward mobility conditional on income segregation, but not vice versa, suggesting spacial segregation has more explanatory power than income segregation;
Descriptive Parameters and Estimates

4. Average CZ income and top 1% share of income are uncorrelated with upward mobility;

5. Gini coefficients have a strong negative correlation with upward mobility, similar to the Great Gatsby Curve, while fraction in middle class has positive correlation;

6. Tax rates, public expenditures, EITC rates, and the difference between top and bottom tax rates are positively correlated with upward mobility;

7. High school expenditures, small class size, high income-adjusted test scores, and low income adjusted dropout rates are correlated with high upward mobility;

8. Number of colleges, average tuition, graduation rate, distance to colleges, number of low-cost colleges, and value of grants to enrolled students are uncorrelated with upward mobility;
Migration inflow and outflow rates as well as the fraction of individuals in the CZ born outside of the CZ are uncorrelated with upward mobility;

Employment rate, fraction of workers employed in the manufacturing industry, and Chinese imports per worker (as a proxy for import competition) are uncorrelated with upward mobility;

Teenage labor force participation is strongly positively correlated with upward mobility;

Social capital and religiosity are strongly positively correlated with upward mobility, while violent crime is strongly negatively correlated;

Fraction of single parents and divorced adults are positively correlated with upward mobility, while fraction married adults is negatively correlated;
What Can We Conclude?

• For the first time, we have population-wide intergenerational administrative data on earnings, family structure, and transfers in the USA;

• We can match the population to commuting zones, and we have much data on characteristics of commuting zones;

• The paper presents interesting descriptive results showing:
  • The rank-rank regression coefficient is around 0.341;
  • The IGE is 0.344, though the IGE is not robust;
  • The rank-rank regression coefficients vary widely depending on region of the country, with the Southeast having the lowest mobility;
  • Various local characteristics that we might expect to be correlated with upward mobility (redistributive policies, lack of segregation, low inequality) are, while others defy expectations (mean income, unemployment, college cost and proximity).

Bradley J. Setzler, University of Chicago

Understanding Chetty, et al. (2014)
What Can We Conclude?

But can we conclude the following?


- “As I said, this observation clearly reinforces the case for policies that help families function without multiple cars.” Paul Krugman, New York Times, July 28, 2013.

- “Why is the American Dream dead in the South?... The top 1 percent aren’t killing the American Dream. Something else is - if you live in the wrong place.” Matthew O’Brien, The Atlantic, January 26, 2014.
What Can We Conclude?

Are the results truly describing the population? The study excludes non-citizens. Here is a 2000 Census map of non-citizens in the USA who entered in 1995-2000.
What Can We Conclude?

Could the lack of relationship between average income and earnings mobility be driven by the choice to use a metric that is not sensitive to levels? Here is 2009 Census median income:

Bradley J. Setzler, University of Chicago
Understanding Chetty, et al. (2014)
What Can We Conclude?

If these regions stifle upward mobility, how do we explain the overall immigration trend toward immobile regions? This is the migration path from the 2000 Census:

[Map image showing net migration flows between California and other states, with numbers indicating the magnitude of migration.]
Again, I emphasize the disconnect between correlation and causation in this study. The study makes no advances in understanding the causes of intergenerational mobility differences by region or regional characteristics, but finds many correlates that merit further study.
References


Appendix: OCG IGE Time-trend

Using the OCG 1973 data, I regress log son income on log parent income for each cohort of fathers in the sample. Unfortunately, the child is asked only for the parents’ household income range rather than the exact value, e.g., household income of parents was between $4,000 and $4,999. I imputed exact values, e.g., $4,500, in order to make the IGE regression possible. I did this in two different ways - the baseline imputation uses the middle of the range, while the alternate imputation uses the lowest point in the range for high ranges and the highest point in the range for low ranges. I run the regression separately for the two standard coding procedures for those with zero (or negative) income: omit them or code them to $1. Finally, I run a version of the regression only on the fraction of the sample within one standard deviation of the mean, thus excluding the outer tails of the distribution. Because so many imputation decisions must be made to arrive at IGE estimates from OCG data, these estimates should not be trusted.

Bradley J. Setzler, University of Chicago

Understanding Chetty, et al. (2014)
Appendix: OCG IGE Time-trend

I find the following time-trend for the IGE of OCG white parent-son sets when those reporting zero (or negative) income are excluded, under various imputations for parent’s income level:

Bradley J. Setzler, University of Chicago
Understanding Chetty, et al. (2014)
I find the following time-trends for the IGE of OCG white parent-son sets when those reporting zero (or negative) income are coded as earning $1, under various imputations for parent’s income level:

Bradley J. Setzler, University of Chicago
Understanding Chetty, et al. (2014)
I find the following time-trends for the IGE of OCG black parent-son sets when those reporting zero (or negative) income are excluded, under various imputations for parent’s income level:

![Graph showing time-trends for the IGE of OCG black parent-son sets.](image-url)
Appendix: OCG IGE Time-trend

I find the following time-trends for the IGE of OCG black parent-son sets when those reporting zero (or negative) income are coded as earning $1, under various imputations for parent’s income level: