



TEMPLE
UNIVERSITY®

**Intergenerational Income Mobility: Counterfactual
Distributions with a Continuous Treatment**

Brantly Callaway
Department of Economics
Temple University

Weige Huang
Department of Economics
Temple University

Department of Economics
DETU Working Paper 18-01
January 2018

1301 Cecil B. Moore Avenue, Philadelphia, PA 19122

<http://www.cla.temple.edu/economics/faculty/detu-working-paper-series/>

Intergenerational Income Mobility: Counterfactual Distributions with a Continuous Treatment*

Brantly Callaway[†]

Weige Huang[‡]

January 23, 2018

Abstract

A vast literature in labor economics documents the correlation between parents' income and their child's income. This paper develops new methods to study the entire distribution of child's income as a function of parents' income while (possibly) adjusting for differences in background characteristics (e.g. race or parents' education) across children whose parents had different incomes. The main challenge is that parents' income is a continuous variable and existing methods for generating counterfactual distributions consider the case with discrete groups. We develop new semiparametric estimators for counterfactual distributions with a continuous treatment variable that are simple to implement and compute quickly. We show that our estimators converge uniformly to Gaussian processes and that the empirical bootstrap can be used to conduct uniformly valid inference across a range of values of parents' income. Our main analysis focuses on parameters such as (i) the fraction of children below the poverty line, (ii) "local" intergenerational elasticities (LIGE), and (iii) the variance of child's income that are functions of the counterfactual distribution. We find variation in these parameters across parents' income levels. And, after documenting large differences in background characteristics across parents' income levels, we find that adjusting for differences in background characteristics tends to reduce (though not eliminate) the overall effect of parents' income on child's income as well as reduce differences in the LIGE and variance of child's income across parents' income levels.

Keywords: Intergenerational Income Mobility, Counterfactual Distribution, Continuous Treatment, Treatment Effects, Distribution Regression

JEL Codes: C21, J62

*We thank Afrouz Azadikhah-Jahromi, David Kaplan, Martin Lopez-Daneri, Catherine Maclean, Pedro Sant'anna, Tymon Sloczynski, Emmanuel Tsyawo, Doug Webber, participants in Temple University's Econometrics Reading Group, and seminar participants at the 2017 Midwest Econometrics Group for valuable comments. Code for our method is available in the R package `ccfa` available on CRAN and Github (<https://github.com/WeigeHuangEcon/ccfa> or `devtools::install_github('WeigeHuangEcon/ccfa')`).

[†]Assistant Professor, Department of Economics, Temple University. Email: brantly.callaway@temple.edu

[‡]PhD Student, Department of Economics, Temple University. Email: weige.huang@temple.edu

1 Introduction

It is well known that children from families with high income tend to have higher incomes than children from families with low income (see Solon (1992) and Solon (1999), among many others). However, much less is known about the distribution of children’s income across parents’ income levels. And learning about this distribution provides much more information to researchers and policy makers about the effect of parents’ income on child’s income.

To give an example, our baseline estimates suggest that a child whose parents’ income is at the poverty line (we set this to be \$22,100 and discuss why below) has an income of \$32,400 on average. If this is all that a researcher knows about outcomes for children from families right at the poverty line, it could be the case that (i) the variance of these individuals’ income is low implying that many of them have incomes very close to \$32,400, or (ii) the variance of these individuals’ income is high implying that some of them have much higher incomes than \$32,400 and others have much lower incomes. In the first case, most children from low income families would be moving out of poverty and into the lower middle class; while in the second case, many children would remain in poverty while others might have substantially higher incomes. These two scenarios have quite different implications for our understanding of the effect of intergenerational income mobility; in particular, if a researcher is interested in the role that parents’ income plays in transmitting poverty, only knowing average income as a function of parents’ income is not enough.

A second motivation of our paper is to look at the role that background characteristics play in the transmission of income across generations. There are many other “covariates” that are also correlated with both parents’ income and child’s income – for example, race and parents’ education. This paper develops new techniques to adjust for differences in covariates across individuals whose parents had different incomes.

From a methodological perspective, the main challenge is that parents’ income is a continuous variable. There is a large literature on estimating counterfactual distributions with discrete groups which includes DiNardo, Fortin, and Lemieux (1996), Machado and Mata (2005), Firpo (2007), Firpo, Fortin, and Lemieux (2009), Rothe (2010), and Chernozhukov, Fernandez-Val, and Melly (2013) among others. One idea would be to divide parents’ income into a small number of groups

and use techniques from this literature. However, this approach would suffer from requiring us to choose cutoffs of parents’ income in some ad hoc way (see Bhattacharya and Mazumder (2011) for similar arguments about the cutoffs required for transition matrices). Instead, we keep parents’ income as a continuous variable and develop new tools to for counterfactual distributions with a continuous treatment.

Because we keep parents’ income as a continuous variable, our method has similarities to the literature on continuous treatment effects (Hirano and Imbens 2004; Flores 2007; Florens, Heckman, Meghir, and Vytlacil 2008; Galvao and Wang 2015; Kennedy, Ma, McHugh, and Small 2016). There are two main differences between our approach and the continuous treatment effect literature. First, we do not impose the condition that the treatment level (or “dose”) is as good as randomly assigned after conditioning on covariates; this assumption is called unconfoundedness, ignorability, or selection on observables. However, our counterfactual distributions correspond exactly to dose-response functions with a continuous treatment under the assumption of unconfoundedness. Within the continuous treatment effect literature, the only paper that we are aware of that looks at distributional parameters is Galvao and Wang (2015) which considers estimation of quantiles with a continuous treatment under unconfoundedness. Our approach is different in that our estimators are based on first step distribution regression and do not require estimating conditional densities in the first step.¹ There are tradeoffs to using distribution regression relative to weighting estimators based on conditional densities. Distribution regression imposes stronger parametric assumptions than nonparametrically estimating conditional densities, though it is much simpler to implement in practice; on the other hand, distribution regression is much more flexible (though perhaps somewhat harder to implement) than assuming a fully parametric model for a conditional density. Thus, our method could be used to estimate the causal effect of a continuous treatment on some outcome of interest under the assumption of unconfoundedness. And the primary difference is one of interpretation; we are careful to use terminology such as

¹Another primary difference between our approach and that of Galvao and Wang (2015) is that quantile dose-response functions are not our primary object of interest. For studying intergenerational mobility, we found that several other parameters (discussed in detail in Section 2) that are functions of the counterfactual distribution are more useful. However, it seems that it would be possible to extend the results in Galvao and Wang (2015) to cover these parameters as well.

“after adjusting for differences in covariates” rather than the language of causal effects.

The counterfactual distribution of child’s incomes is built by *fixing* the distribution of child’s income conditional on parents’ income and observed characteristics but *changing* the distribution of observed characteristics conditional on parents’ income. In particular, we consider changing the distribution of observed covariates conditional on parents’ income to be the distribution of covariates for the entire population. To give an example, suppose the only covariate is parents’ education and that parents’ education is positively related to parents’ income and child’s income. Further, suppose that we are interested in the distribution of child’s income conditional on parents having low income. To obtain a counterfactual distribution, we fix the distribution of child’s income conditional on both education and parents’ income, but change the distribution of education to be that of the entire population – thus putting relatively more weight on the income of children with highly educated parents who had low income.

The resulting counterfactual distribution is difficult to work with and not easy to directly understand because it is a function of both child’s income and parents’ income. Instead, we focus on various functionals of the counterfactual distribution that are functions only of parents’ income. In particular, we consider (1) average child’s income, (2) the “local” intergenerational elasticity, (3) the variance of child’s income, (4) the fraction of children whose incomes are below the poverty line, and several others – all as a function of parents’ income. Each of these parameters can be plotted in two dimensions and the results are easy to interpret.

Our estimates should not, in general, be interpreted as the causal effect of parents’ income on child’s income. As mentioned above, under the assumption of unconfoundedness, our counterfactual distributions correspond to the dose-response function of child’s income conditional on parents’ income – an object of central interest in the literature on continuous treatment effects; however, we arrive at this counterfactual distribution as a result of manipulating the distribution of observed characteristics, not from defining potential outcomes and assuming unconfoundedness.

In order to estimate counterfactual distributions in this framework, we develop new semiparametric estimators of counterfactual intergenerational mobility parameters while treating parents’ income as being continuous. Estimation requires three steps. First, we estimate the distribution

of child's income conditional on parents' income and observed characteristics using distribution regression (Foresi and Peracchi 1995; Chernozhukov, Fernandez-Val, and Melly 2013). The second step is to estimate the counterfactual distribution which involves integrating over the first step estimates while changing the distribution of observable characteristics. Finally, parameters of interest such as measures of the spread of child's income and the probability of child's income being below the poverty line are obtained as functions of the counterfactual distribution. We obtain the limiting processes for each of our parameters of interest and develop inference procedures using results from the empirical process literature (see, for example, Van Der Vaart and Wellner (1996) and Kosorok (2007)). We show that the limiting processes can be approximated using the empirical bootstrap. These results allow us to test functional hypotheses about income mobility such as (1) whether adjusting for covariates has any effect on any particular parameter (e.g. local measures of the intergenerational elasticity or the percentage of children with income below the poverty line as a function of parents' income), (2) whether parameters of interest are the same at all values of parents' income (e.g. the variance of child's income), among others.

Like most of the intergenerational income mobility literature, we find a strong relationship between parents' income and child's income. Without adjusting for any differences in covariates, we find that (1) children from low income families have lower income on average than children from high income families; (2) children from low income families have on average higher income than their parents, while children from high income families have lower income on average than their parents; (3) children from low income families are much more likely to have income below the poverty line than children from high income families; (4) children from low income families are much less likely to be in the top 10% of income than children from high income families; (5) children from low income families may have somewhat higher variance in their earnings than children from high income families. The first two of these results are in common with almost all of the intergenerational mobility literature. The latter three are new results and available because of the new methods developed in the paper.

We also find that background characteristics such as parents' education, race, and whether or not a child is from a single parent household, are strongly correlated with parents' income. We

find that adjusting for covariates does not overturn any of the five main results above; however, overall adjusting for differences in observed characteristics across parents' income levels does tend to reduce the effects of parents' income. Adjusting for differences in observed characteristics flattens somewhat the relationship between child's income and parents' income. It also reduces by about one third the estimated probability that a child's income will be below the poverty rate for children from families with income close to the poverty line. Taken together, our results suggest that differences in background characteristics explain some, but not all, of the differences in outcomes experienced by children whose parents had different incomes.

Our results are related to work that has used quantile regression to study intergenerational mobility (Eide and Showalter 1999; Grawe 2004). These papers show that the distribution of child's income conditional on parents' income narrows as parents' income increases. Our unconditional results can be compared directly with the results in those papers. However, our counterfactuals are fundamentally different than quantile regression specifications that include additional control variables.

Finally, the literature on intergenerational income mobility is vast and we briefly summarize some of the most relevant parts (a much more detailed review of the literature can be found in Black and Devereux (2011)). Regressing the log of child's income on the log of parents' income has a long history in the intergenerational income mobility literature – the resulting estimate is called the intergenerational elasticity (IGE). But recent work has considered more complicated setups such as (1) transition matrices,² (2) the probability that child's income is greater than parents' income, and (3) the correlation of the ranks of child's income and parents' income (Jantti et al. 2006; Bhattacharya and Mazumder 2011; Chetty, Hendren, Kline, and Saez 2014; Chetty et al. 2014; Chetty et al. 2017; Collins and Wanamaker 2017). Bowles and Gintis (2002), Groves

²Transition matrices have received considerable attention in the intergenerational income mobility literature (see Jantti et al. (2006), Bhattacharya and Mazumder (2011), Black and Devereux (2011), and Richey and Rosburg (2015), among others). In principle, one could use a transition matrix to calculate the probability that a child's income is below the poverty line for different values of parents' income. However, transition matrices typically pick cutoff points at particular quantiles of parents' income (e.g. at the 25th, 50th, and 75th percentiles) and look at quantiles of child's income as well. This turns out to be an important distinction. Because quantiles of income depend both on an individual's income and on the income of other individuals, transition matrices are relative mobility measures. On the other hand, calculating the probability that a child's income is below the poverty line as a function of parents' income is an absolute mobility measure as it does not depend on outcomes for other individuals.

(2005), Blanden, Gregg, and Macmillan (2007), and Richey and Rosburg (2017) have decomposed intergenerational mobility parameters into parts that are explained by various background characteristics. There has also been work that tries to identify the causal effect of parents’ income on child’s income from natural experiments such as adoptions, changes in tax policies, and job displacements (Björklund, Lindahl, and Plug 2006; Bratberg, Nilsen, and Vaage 2008; Oreopoulos, Page, and Stevens 2008; Liu and Zeng 2009; Currie and Almond 2011; Dahl and Lochner 2012).

Terminology The data that we use consists of measures of parents’ permanent income and their child’s permanent income. We use the shorthand terminology *parents’ income* for the total family permanent income of the parents and the shorthand *child’s income* for the total family permanent income of the child. Some variables are computed for the “head” of the family which refers to one particular parent, usually the parent with the higher income; sometimes the head of the family changes – in these cases, we consider the parent who is listed as the head for the most years while a child is a member of the household to be the household head.

Notation Y indicates the log of child’s income. T is the log of parents’ income. X indicates other covariates that we adjust for in creating counterfactual distributions. $F_{Y|T}$ is the distribution of child’s income conditional on parents’ income. We refer to this distribution as the “observed distribution” throughout the paper. $F_{Y|T}^C$ is the counterfactual distribution of child’s income conditional on parents’ income. We refer to this distribution as the “counterfactual distribution” throughout the paper.

2 Parameters of Interest

Our approach is different from existing work on intergenerational income mobility in three ways. First, we keep parents’ income as a continuous variable and all of our results are “local”; that is, conditional on a particular value of parents’ income. This setup is different from most work on intergenerational mobility that either estimates a single intergenerational mobility parameter or breaks the observations into several groups. Second, our method allows us to look at the

entire distribution of child’s income conditional on parents’ income. This allows us to estimate parameters such as the fraction of children below the poverty line or the variance of child’s income, both as a function of parents’ income. These parameters provide much more information about outcomes of children given their parents’ income than simply computing the average. Finally, we also are interested in comparing these parameters that can be obtained directly from the observed data to ones that result from “adjusting” the effect of parents’ income for differences in observable characteristics. This section details the ideas behind each of these three contributions. Our starting point is that we have a sample of observations from the joint distribution (Y, T, X) where Y is log of child’s income, T is log of parents’ income and X are additional covariates such as parents’ education, child’s birth year, gender, and race.

2.1 Counterfactuals

The first step in our analysis is to study how much intergenerational income mobility measures are affected by differences in observed characteristics between parents with low incomes and parents with high incomes. We take as the primary building block of counterfactuals the distribution of child’s income conditional on parents’ income and observed characteristics: $F_{Y|T,X}(y|t, x)$. Throughout, distributions with a C superscript, that is F^C , are counterfactual distributions while distributions without a superscript such as F denote distributions that are identified by the sampling process. Notice that the observed distribution of child’s income conditional on parents’ income is given by:

$$F_{Y|T}(y|t) = \int_{\mathcal{X}} F_{Y|T,X}(y|t, x) \, dF_{X|T}(x|t) \quad (2.1)$$

that is, $F_{Y|T}(y|t)$ is the same as integrating the distribution conditional on observed characteristics X over the distribution of X conditional on $T = t$. The counterfactuals that we consider hold $F_{Y|T,X}(y|t, x)$ constant in Equation 2.1 while changing the distribution of other characteristics X conditional on parents’ income T . For a fixed parents’ income t , the counterfactual distribution that we consider adjusts the distribution of covariates X conditional on $T = t$ to be given by the

distribution of X in the population. To give an example, let t be relatively small and consider the case where parents' education is positively correlated with parents' income. Then, the counterfactual distribution would put more weight on observations with $T = t$ that have high levels of education and less weight on observations with $T = t$ that have low levels of education. This counterfactual distribution is given by

$$F_{Y|T}^C(y|t) = \int_{\mathcal{X}} F_{Y|T,X}(y|t, x) \, dF_X(x) \quad (2.2)$$

This particular counterfactual is exactly the same as if one were to impose unconfoundedness and calculate the distribution of “potential” outcomes. Although it is not equal to the observed distribution, all the terms on the right hand side are identified and one can estimate this counterfactual distribution by plugging in to the above equation. While it is possible to show that Equation 2.2 is equivalent to a weighting estimator (weighting estimators are developed in DiNardo, Fortin, and Lemieux (1996) and Firpo (2007) in the case where the treatment is binary and in Galvao and Wang (2015) in the continuous treatment case), we find it more natural to estimate the conditional distribution directly in Equation 2.2 which is more similar to the approaches taken in Machado and Mata (2005), Melly (2005), and Chernozhukov, Fernandez-Val, and Melly (2013), all in the case where the treatment is a discrete variable. The reason is that, with a continuous treatment variable, the weights are given by conditional density functions³ which are more challenging to estimate than the conditional distribution function above.

2.2 Parameters of Interest

The observed distribution $F_{Y|T}$ of child's income conditional on parents' income and the counterfactual distribution $F_{Y|T}^C$ contain much information about intergenerational mobility, but they suffer from being difficult to interpret or plot directly. Instead, we focus on estimating functionals of $F_{Y|T}$ and $F_{Y|T}^C$. This section covers these functionals.

³With a binary (or even discrete) treatment, the weights depend on the propensity score which is much more straightforward to estimate – for example, one could use logit or probit.

Fraction of Children in Poverty and “Rich” Children Conditional on Parents’ Income

The first parameter that we consider is the fraction of children whose permanent income falls below the poverty line as a function of parents’ income. Let y_p denote the poverty line. Then, the fraction of children below the poverty line is given by

$$F_{Y|T}(y_p|t) \quad \text{and} \quad F_{Y|T}^C(y_p|t)$$

for the fraction below the poverty line coming from the observed data and from the counterfactual distribution, respectively. These are straightforward measures to plot, as a function of t , and if children with lower income parents are more likely to have permanent incomes below the poverty line, then one would expect that this line would be downward sloping.

To examine the roll that adjusting for differences in covariates plays in terms of the fraction of children with income below the poverty line, we also consider the parameter

$$\Delta^{POV}(t) = F_{Y|T}(y_p|t) - F_{Y|T}^C(y_p|t)$$

which is the difference in the poverty rates coming from the observed distribution and the counterfactual distribution for some particular value of parents’ income t .

We are also interested in the fraction of children that have very high permanent income. Let y_R be some particular value of child’s permanent income – later we set this to be the 90th percentile of income in the U.S. in 2010. Then, the fraction of “rich” children conditional on parents’ income is given by

$$1 - F_{Y|T}(y_R|t) \quad \text{and} \quad 1 - F_{Y|T}^C(y_R|t)$$

coming from the observed distribution and the counterfactual distribution, respectively. Like the fraction below the poverty line, these are straightforward to plot and interpret, and we can consider

the parameter

$$\Delta^{RICH}(t) = (1 - F_{Y|T}(y_R|t)) - (1 - F_{Y|T}^C(y_R|t)) = F_{Y|T}^C(y_R|t) - F_{Y|T}(y_R|t)$$

which is the difference in the fraction of children that are rich coming from the observed distribution relative to the counterfactual distribution.

Quantiles of Child's Income Conditional on Parents' Income

One can obtain the quantiles of child's income conditional on parents' income from the observed distribution and counterfactual distributions. For some $\tau \in (0, 1)$, these are given by

$$Q_{Y|T}(\tau|t) = \inf\{y : F_{Y|T}(y) \geq \tau\} \quad \text{and} \quad Q_{Y|T}^C(\tau|t) = \inf\{y : F_{Y|T}^C(y) \geq \tau\}$$

for the conditional quantiles of the observed distribution and the conditional quantiles of the counterfactual distribution, respectively. $Q_{Y|T}^C$ is called the quantile dose-response function in Galvao and Wang (2015). The quantiles are also useful inputs into the remaining parameters of interest. The difference between estimates coming from the observed distribution and counterfactual distribution is given by

$$\Delta^Q(\tau|t) = Q_{Y|T}(\tau|t) - Q_{Y|T}^C(\tau|t)$$

and is also a parameter of interest.

Average Child's Income Conditional on Parents' Income

The next parameter that we consider is average child's income conditional on parents' income which is given by

$$E[Y|T = t] = \int_0^1 Q_{Y|T}(\tau|t) \, d\tau \quad \text{and} \quad E^C[Y|T = t] = \int_0^1 Q_{Y|T}^C(\tau|t) \, d\tau$$

where these depend on the observed distribution and counterfactual distribution, respectively. To

assess the effect of covariates on average child's income conditional on parents' income, one can also consider the parameter

$$\Delta^E(t) = E[Y|T = t] - E^C[Y|T = t]$$

For some value $t \in \mathcal{T}$, $\Delta^E(t) > 0$ implies that adjusting for covariates lowers average income for children with parents' with income t . If covariates, such as education, are positively related to parents' income and positively related to child's income, then one would expect that $\Delta^E(t)$ would be negative for small values of t and positive for large values of t .

Average child's income conditional on parents' income is closely related to the Intergenerational Elasticity (IGE) that is very commonly estimated in the intergenerational mobility literature. IGE is the coefficient on the log of parents' income in the regression of log child's income on log parents' income. The slope of $E[Y|T = t]$ corresponds to the IGE though, in our case, the slope is not restricted to be constant. Our measure is therefore more similar to the Local IGE (LIGE) measure in Landersø and Heckman (2017) (see also Bratsberg et al. (2007) and Björklund, Roine, and Waldenström (2012) for other local versions of the IGE). In addition, the slope of $E^C[Y|T = t]$ is a local measure of IGE, after adjusting for differences in covariates at different values of t . We consider the local intergenerational elasticity given by

$$LIGE(t) = \frac{E[Y|T = t + \delta] - E[Y|T = t - \delta]}{2\delta}$$

and

$$LIGE^C(t) = \frac{E^C[Y|T = t + \delta] - E^C[Y|T = t - \delta]}{2\delta}$$

where δ is some small fixed number. We also consider

$$\Delta^{LIGE}(t) = LIGE(t) - LIGE^C(t)$$

which is the difference between LIGE estimates coming from the observed distribution and coun-

terfactual distribution.

Measures of Spread of Child's Income Conditional on Parents' Income

Looking at average child's income conditional on parents' income is very similar to most existing work on intergenerational mobility. However, because our method obtains the entire observed distribution and counterfactual distribution of child's income conditional on parents' income, we can study other features of these distributions than just their mean. In this section, we consider the variance of child's income conditional on parents' income and the inter-quantile range of child's income conditional on parents' income. These give measures of the spread of child's income conditional on parents' incomes.

Given the existing results in the intergenerational mobility literature, one would strongly suspect that child's income tends to increase with parents' income, at least on average. However, much less is known about the spread of child's income conditional on parents' income. It is possible that the distribution of child's income simply shifts to the right as parents' income increases. If the variance of child's income decreases with parents' income, that would suggest that having parents with high income increases income on average and increases the certainty of obtaining higher income. Decreasing variance would also suggest that the income of children from low income families is riskier. On the other hand, if the variance of child's income is increasing in parents' income, that would suggest that children from high income families are more likely to become very rich but also have some risk of having low incomes (and the reverse would be true for children of low income families).

The first measure of spread that we consider is the variance of child's income conditional on parents' income. It is given by

$$Var(Y|T = t) = \int_0^1 (Q_{Y|T}(\tau|t) - E[Y|T = t])^2 d\tau$$

and

$$Var^C(Y|T = t) = \int_0^1 (Q_{Y|T}^C(\tau|t) - E^C[Y|T = t])^2 d\tau$$

The second measure of spread is an inter-quantile range which is given by

$$IQR(\tau_1, \tau_2; t) = Q_{Y|T}(\tau_1|t) - Q_{Y|T}(\tau_2|t) \quad \text{and} \quad IQR^C(\tau_1, \tau_2; t) = Q_{Y|T}^C(\tau_1|t) - Q_{Y|T}^C(\tau_2|t)$$

where $\tau_1 > \tau_2$. A typical example would be to look at the spread between the 90th percentile of child's income and 10th percentile of child's income conditional on parents' income being given by t . We also define the differences between our estimators of the spread of income conditional on parents' income from the observed distribution and counterfactual distributions by

$$\Delta^{Var}(t) = Var(Y|T=t) - Var^C(Y|T=t) \quad \text{and} \quad \Delta^{IQR}(\tau_1, \tau_2; t) = IQR(\tau_1, \tau_2; t) - IQR^C(\tau_1, \tau_2; t)$$

2.3 Testing if Parameters Depend on Parents' Income

Each of the parameters mentioned above can be considered as a function of parents' income t . As a final step in our analysis, we are interested in testing whether parents' income has any effect on the parameters of interest. Let $\theta(t)$ denote a generic parameter of interest – this includes parameters obtained from the observed distribution or the counterfactual distribution. Then, we are interested in the null hypothesis that

$$\theta(t) = E[\theta(T)] \quad \text{for all } t \in \mathcal{T}$$

Let $R_\theta(t) = \theta(t) - E[\theta(T)]$. We are interested in testing the following hypothesis

$$H_0 : R_\theta(t) = 0 \quad \text{for all } t \in \mathcal{T} \tag{2.3}$$

To give an example, one could be interested in testing whether the variance of child's income changes with parents' income, both using the observed distribution and using the counterfactual distribution that adjusts for differences in the distribution of covariates across different levels of parents' incomes. This sort of test allows one to do exactly that.

3 Estimation

Estimation proceeds in three steps. In step 1, we estimate the distribution of Y (log child's income) conditional on T (log parents' income) and possibly X (other observed characteristics) using distribution regression. For counterfactual distributions, step 2 involves integrating the conditional distribution over a counterfactual distribution of X conditional on T . In particular, we consider the counterfactual distribution $F_{X|T}^C = F_X$; that is, we set the distribution of X conditional on T to be equal to the distribution of X for the overall population for all values of T . With step 2 complete, we have a (counterfactual) distribution of Y conditional on T . The final step is to manipulate the (counterfactual) distribution into the particular parameters of interest given in Section 2. Throughout, we assume that we have a random sample of units, indexed by $i = 1, \dots, n$. We observe dependent variable Y_i , treatment variable T_i and a vector of covariates X_i for each unit i . We assume the dimension of X is k and includes a constant.

3.1 Step 1: Estimating the Conditional Distribution

We estimate the conditional function $F_{Y|T,X}$ using distribution regression (Foresi and Peracchi 1995; Chernozhukov, Fernandez-Val, and Melly 2013). The main idea here is to estimate a series of binary response models using $\mathbb{1}\{Y \leq y\}$ as the dependent variable while varying y ;⁴ that is

$$\begin{aligned} F_{Y|X,T}(y|x, t) &= E[\mathbb{1}\{Y \leq y\} | X = x, T = t] \\ &= \Lambda(\alpha_1(y)t + x'\alpha_2(y)) \end{aligned}$$

where Λ is a known link function – we use the logistic link function though one could make some other choice here. $\alpha_1(y)$ and $\alpha_2(y)$ are unknown parameters corresponding to each y , i.e., the parameters α_1 and α_2 change as y changes. $\mathbb{1}\{Y \leq y\}$ is an indicator function that equals one if

⁴A reasonable alternative approach would be to use quantile regression (Koenker and Bassett Jr 1978; Koenker 2005) in the first step to obtain estimates of conditional quantiles, then invert them to obtain conditional distributions. From there, all the other steps would be exactly the same. Our approach models the conditional distributions directly using distribution regression.

$Y \leq y$ is true and zero otherwise.

To implement the distribution regression estimator, we estimate a series of logit models over a fine grid of possible values for y . The estimated conditional distribution is⁵

$$\hat{F}_{Y|X,T}(y|x,t) = \Lambda(\hat{\alpha}_1(y)t + x'\hat{\alpha}_2(y)) \quad (3.1)$$

We plug these estimators in to the counterfactual operations discussed next.

Similarly, we also estimate $F_{Y|T}(y|t)$ (the observed distribution of child's income conditional on parents' income) using distribution regression. Here, we suppose that

$$F_{Y|T}(y|t) = \Lambda(\beta_0(y) + \beta_1(y)t)$$

and estimate the parameters $\beta_0(y)$ and $\beta_1(y)$ using logit over a fine grid of values for y . Then, the estimated value of $F_{Y|T}(y|t)$ is given by

$$\hat{F}_{Y|T}(y|t) = \Lambda(\hat{\beta}_0(y) + \hat{\beta}_1(y)t)$$

3.2 Step 2: Estimating Counterfactual Distributions

From the subsection above, we obtained an estimator of $F_{Y|T,X}$. For fixed y and t , estimating $F_{Y|T}^C(y|t)$ amounts to averaging over X while holding t fixed. That is,

$$\begin{aligned} \hat{F}_{Y|T}^C(y|t) &= \frac{1}{n} \sum_{i=1}^n \hat{F}_{Y|T,X}(y|t, X_i) \\ &= \frac{1}{n} \sum_{i=1}^n \Lambda(\hat{\alpha}_1(y)t + X_i'\hat{\alpha}_2(y)) \end{aligned} \quad (3.2)$$

which is the same as replacing the population distribution function in Equation 2.2 with the sample distribution function. We plug in the estimates $\hat{F}_{Y|T}$ and $\hat{F}_{Y|T}^C$ below to obtain estimates

⁵Since the estimated conditional distribution obtained above may be nonmonotonic in y , we apply the monotization method of Chernozhukov, Fernández-Val, and Galichon 2010 based on rearrangement. In practice, we use rearranged estimators of the conditional distribution for all the results, but we omit this discussion throughout the rest of this section for the sake of clarity.

of particular parameters of interest.

3.3 Step 3: Estimating Parameters of Interest

Once the observed distribution and counterfactual distribution of child's income conditional on parents' income have been estimated, one can estimate the parameters of interest considered in Section 2. Estimating the fraction of children with income below the poverty line is straightforward and given by

$$\hat{F}_{Y|T}(y_p|t) \quad \text{and} \quad \hat{F}_{Y|T}^C(y_p|t)$$

Estimating quantiles of child's income conditional on parents' income can also be obtained simply by plugging in to the results in Section 2:

$$\hat{Q}_{Y|T}(\tau|t) = \inf\{y : \hat{F}_{Y|T}(y|t) \geq \tau\} \quad \text{and} \quad \hat{Q}_{Y|T}^C(\tau|t) = \inf\{y : \hat{F}_{Y|T}^C(y|t) \geq \tau\}$$

which simply inverts the counterfactual distribution of outcomes. Next, we consider estimating $E[Y|T = t]$. Consider a grid of equally spaced values of τ given by $0 < \tau_1 < \tau_2 < \dots < \tau_S < 1$. Then, we estimate $E[Y|T = t]$ by

$$\hat{E}[Y|T = t] = \frac{1}{S} \sum_{s=1}^S \hat{Q}_{Y|T}(\tau_s|t) \quad \text{and} \quad \hat{E}^C[Y|T = t] = \frac{1}{S} \sum_{s=1}^S \hat{Q}_{Y|T}^C(\tau_s|t)$$

where S is some large positive integer. These estimates can be plugged in to obtain estimates of the LIGE given by

$$LIGE(t) = \frac{\hat{E}[Y|T = t + \delta] - \hat{E}[Y|T = t - \delta]}{2\delta}$$

and

$$LIGE^C(t) = \frac{\hat{E}^C[Y|T = t + \delta] - \hat{E}^C[Y|T = t - \delta]}{2\delta}$$

for some fixed δ . We can also estimate the conditional variance by plugging in

$$\hat{Var}(Y|T=t) = \frac{1}{S} \sum_{s=1}^S \left(\hat{Q}_{Y|T}(\tau_s|t) - \hat{E}[Y|T=t] \right)^2$$

and

$$\hat{Var}^C(Y|T=t) = \frac{1}{S} \sum_{s=1}^S \left(\hat{Q}_{Y|T}^C(\tau_s|t) - \hat{E}^C[Y|T=t] \right)^2$$

Finally, estimates of the inter-quantile range are given by

$$I\hat{Q}R(\tau_1, \tau_2; t) = \hat{Q}_{Y|T}(\tau_1|t) - \hat{Q}_{Y|T}(\tau_2|t) \quad \text{and} \quad I\hat{Q}R^C(\tau_1, \tau_2; t) = \hat{Q}_{Y|T}^C(\tau_1|t) - \hat{Q}_{Y|T}^C(\tau_2|t)$$

4 Asymptotic Theory

This section develops asymptotic theory and inference procedures for the parameters discussed in Section 2. Our inference results are uniformly valid in parents' income T . And we show that each of the parameters that we consider converges uniformly to a Gaussian process. These results allow us to test functional hypotheses such as (1) whether the results from adjusting for differences in other covariates X are different from the results obtained directly from the observed data at *any* value of parents' income, (2) whether any parameter of interest (such as the variance or inter-quantile range) of child's income is constant across parents' income levels, among others. We develop these asymptotic results using arguments from the empirical processes literature (see, for example Van Der Vaart and Wellner 1996; Kosorok 2007). For any discrete set of values of T , a Gaussian process is just a (multivariate) normal distribution, so our results also contain as special cases pointwise results. The second part of the results in this section shows that the empirical bootstrap is valid for conducting inference – both uniformly and pointwise. All proofs are contained in the appendix. We make the following assumption,

Assumption 1. (*Random Sampling*)

$\{Y_i, T_i, X_i\}_{i=1}^n$ are iid draws from the joint distribution $F_{Y,T,X}$.

Several other standard assumptions for Distribution Regression and other technical conditions are collected in Assumption A.1 in the appendix. We use the following notation. Let $l^\infty(S)$ denote the space of all uniformly bounded functions on the set S equipped with the supremum norm denoted $\|\cdot\|_\infty$. Let \mathcal{Y} , \mathcal{T} , and \mathcal{X} denote the supports of Y , T , and X , respectively. Let

$$\hat{G}_{Y|T}^C(y|t) = \sqrt{n}(\hat{F}_{Y|T}^C(y|t) - F_{Y|T}^C(y|t))$$

denote the empirical process of the counterfactual distribution of child's income conditional on parents' income. Further, let

$$\hat{G}_{Y|T}(y|t) = \sqrt{n}(\hat{F}_{Y|T}(y|t) - F_{Y|T}(y|t))$$

denote the empirical process of the observed distribution of child's income conditional on parents' income.

Theorem 1 establishes the joint limiting process for the observed distribution and counterfactual distribution.

Theorem 1. *Let $\mathbb{S} = l^\infty(\mathcal{Y}\mathcal{T})^2$. Under Assumption 1 and A.1 (given in Appendix A)*

$$(\hat{G}_{Y|T}(y|t), \hat{G}_{Y|T}^C(y|t)) \rightsquigarrow (\mathbb{V}_{Y|T}(y|t), \mathbb{V}_{Y|T}^C(y|t))$$

in the space \mathbb{S} where $(\mathbb{V}_{Y|T}^C, \mathbb{V}_{Y|T})$ is a tight Gaussian process indexed by (y, t) with mean 0 and where $\mathbb{V}_{Y|T}(y|t) = \mathbb{G}_{Y|T}(y|t)$ and $\mathbb{V}_{Y|T}^C(y|t) = \int_{\mathcal{X}} \mathbb{G}_{Y|T,X}(y|t, x) dF_X(x) + \int_{\mathcal{X}} F_{Y|T,X}(y|t, x) d\mathbb{G}_X(x)$ where $\mathbb{G}_{Y|T}$, $\mathbb{G}_{Y|T,X}$, and \mathbb{G}_X are given in Lemma 1 in the appendix.

Theorem 1 is an important building block for establishing the limiting processes of each of the parameters of interest in Section 2. We will show next that each of the parameters of interest is a Hadamard differentiable function of either the counterfactual distribution or the observed distribution. Theorem 1 is also useful because it allows one to consider uniform inference on the *difference* between particular parameters under the observed distribution and counterfactual distributions and will be important for testing whether or not a particular parameter changes as

parents' income changes.

The next theorem establishes the limiting process for the fraction of children who have permanent income below the poverty line as a function of parents' income.

Theorem 2. *Let $\hat{G}_T^{POV}(y_p|t) = \sqrt{n}(\hat{F}_{Y|T}(y_p|t) - F_{Y|T}(y_p|t))$ where $y_p \in \mathcal{Y}$ denotes the poverty line and is fixed; let $\hat{G}_T^{C,POV}(y_p|t) = \sqrt{n}(\hat{F}_{Y|T}^C(y_p|t) - F_{Y|T}^C(y_p|t))$. Under Assumptions 1 and A.1,*

$$(\hat{G}_T^{POV}(y_p|t), \hat{G}_T^{C,POV}(y_p|t)) \rightsquigarrow (\mathbb{V}_{y_p}^{POV}(t), \mathbb{V}_{y_p}^{C,POV}(t))$$

where $\mathbb{V}^{POV}(t)$ is a stochastic process in the metric space $l^\infty(\mathcal{T})$ given by $\mathbb{V}_{Y|T}(y_p|t)$ and where $\mathbb{V}^{C,POV}(t)$ is a stochastic process in the metric space $l^\infty(\mathcal{T})$ given by $\mathbb{V}_{Y|T}^C(y_p|t)$. In addition,

$$\sqrt{n}(\hat{\Delta}^{POV}(y_p, t) - \Delta^{POV}(y_p, t)) \rightsquigarrow \mathbb{V}_{y_p}^{POV}(t) - \mathbb{V}_{y_p}^{C,POV}(t)$$

in the space $l^\infty(\mathcal{T})$.

The results of Theorem 2 apply to the case where the fraction below the poverty line is computed using the observed conditional distribution, the counterfactual distribution, and also provides the limiting process for their difference. This last result allows one to formally test whether accounting for differences in covariates across parents' income accounts for differences in the fraction of children below the poverty line at any level of parents' income.

The next result shows that the observed quantiles of child's income conditional on parents' income and the counterfactual quantiles of child's income conditional on family income converge uniformly both in the quantiles and parents' income to a Gaussian process.

Theorem 3. *Let $\hat{Z}_{Y|T}(\tau|t) = \sqrt{n}(\hat{Q}_{Y|T}(\tau|t) - Q_{Y|T}(\tau|t))$ and let $\hat{Z}_{Y|T}^C(\tau|t) = \sqrt{n}(\hat{Q}_{Y|T}^C(\tau|t) - Q_{Y|T}^C(\tau|t))$. Let $\mathbb{S} = l^\infty(\mathcal{UT})^2$. Under Assumptions 1 and A.1,*

$$(\hat{Z}_{Y|T}(\tau|t), \hat{Z}_{Y|T}^C(\tau|t)) \rightsquigarrow (\mathbb{Z}(\tau|t), \mathbb{Z}^C(\tau|t))$$

in the space \mathbb{S} with

$$\mathbb{Z}(\tau|t) = \frac{\mathbb{V}_{Y|T}(Q_{Y|T}(\tau|t))}{f_{Y|T}(Q_{Y|T}(\tau|t)|t)} \quad \text{and} \quad \mathbb{Z}^C(\tau|t) = \frac{\mathbb{V}_{Y|T}^C(Q_{Y|T}^C(\tau|t))}{f_{Y|T}^C(Q_{Y|T}^C(\tau|t)|t)}$$

Moreover,

$$\sqrt{n}(\hat{\Delta}^Q(\tau|t) - \Delta^Q(\tau|t)) \rightsquigarrow \mathbb{Z}(\tau|t) - \mathbb{Z}^C(\tau|t)$$

in the space $l^\infty(\mathcal{T})$.

Quantiles of counterfactual distributions may be of interest in themselves. We also use Theorem 3 in the next set of theorems. Theorem 3 is also useful as a building block for other parameters of interest, in particular for $E[Y|T = t]$, $Var(Y|T = t)$, and $IQR(\tau_1, \tau_2; t)$ as well as their counterfactual counterparts.

The next theorem establishes the limiting process for average child's income conditional on parents' income which holds uniformly in parents' income.

Theorem 4. Let $\hat{G}_T^E(t) = \sqrt{n}(\hat{E}[Y|T = t] - E[Y|T = t])$ and let $\hat{G}_T^{C,E}(t) = \sqrt{n}(\hat{E}^C[Y|T = t] - E^C[Y|T = t])$. Under Assumptions 1 and A.1,

$$(\hat{G}_T^E(t), \hat{G}_T^{C,E}(t)) \rightsquigarrow (\mathbb{V}_T^E(t), \mathbb{V}_T^{C,E}(t))$$

in the space $l^\infty(\mathcal{T})^2$ where \mathbb{V}_T^E is a tight Gaussian process with mean 0 given by $\mathbb{V}_T^E = \int_0^1 \mathbb{Z}(\tau|\cdot) \, d\tau$ and $\mathbb{V}_T^{C,E}$ is a tight Gaussian process with mean 0 given by $\mathbb{V}_T^{C,E} = \int_0^1 \mathbb{Z}^C(\tau|\cdot) \, d\tau$. In addition,

$$\sqrt{n}(\hat{\Delta}^E(t) - \Delta^E(t)) \rightsquigarrow \mathbb{V}_T^E(t) - \mathbb{V}_T^{C,E}(t)$$

in the space $l^\infty(\mathcal{T})$.

Theorem 4 provides a way to construct uniform confidence bands for $E[Y|T = t]$, $E^C[Y|T = t]$, and their difference. These results allow us to compare average child's income across parents' income and learn about the role that covariates play in the intergenerational transmission of

income.

The final result of this section provides a way to construct uniform confidence bands for the Local Intergenerational Elasticity.

Theorem 5. *Let $\hat{G}_T^{LIGE}(t) = \sqrt{n}(\hat{LIGE}(t) - LIGE(t)$, let $\hat{G}_T^{C,LIGE}(t) = \sqrt{n}(\hat{LIGE}^C(t) - LIGE^C(t))$, and let $\mathcal{T}_\delta = [t_{min} + \delta, t_{max} - \delta]$ where t_{min} and t_{max} are the endpoints of \mathcal{T} (See Assumption A.1(iii)). Under Assumptions 1 and A.1,*

$$(\hat{G}_T^{LIGE}(t), \hat{G}_T^{C,LIGE}(t)) \rightsquigarrow (\mathbb{V}_T^{LIGE}(t), \mathbb{V}_T^{C,LIGE}(t))$$

in the space $l^\infty(\mathcal{T}_\delta)^2$ where \mathbb{V}_T^{LIGE} and $\mathbb{V}_T^{C,LIGE}$ are tight Gaussian processes with mean 0 given by

$$\mathbb{V}_T^{LIGE}(t) = \frac{\mathbb{V}_T^E(t + \delta) - \mathbb{V}_T^E(t - \delta)}{2\delta} \quad \text{and} \quad \mathbb{V}_T^{C,LIGE}(t) = \frac{\mathbb{V}_T^{C,E}(t + \delta) - \mathbb{V}_T^{C,E}(t - \delta)}{2\delta}$$

where \mathbb{V}_T^E and $\mathbb{V}_T^{C,E}$ are given in Theorem 4. In addition,

$$\sqrt{n}(\hat{\Delta}^{LIGE}(t) - \Delta^{LIGE}(t)) \rightsquigarrow \mathbb{V}_T^{LIGE}(t) - \mathbb{V}_T^{C,LIGE}(t)$$

in the space $l^\infty(\mathcal{T}_\delta)$.

Appendix B contains similar results that establish the limiting process for (i) the variance of child's income conditional on parents' income and (ii) the inter-quantile range of child's income conditional on parents' income. These results follow the same pattern as the results developed in this section and are omitted from the current section to save space. Suffice it to say that the variance of child's income and the inter-quantile range, both as functions of parents' income, converge uniformly to Gaussian processes.

4.1 Inference using the Bootstrap

The limiting processes above depend on unknown nuisance parameters which complicate inference. Thus, to conduct inference, we use the empirical bootstrap. This section shows that the

empirical bootstrap procedure can be used to construct asymptotically valid uniform bands for each of the parameters considered above.

Let $\theta(t)$ generically denote one of the parameters of interest in the preceding sections, for example, $F_{Y|T}(y_p|t)$ or $E^C[Y|T = t]$. Let $\hat{\theta}(t)$ denote an estimator of $\theta(t)$. Let $\hat{\theta}^*(t)$ denote a bootstrapped version of the estimator; in other words, computed using draws from the empirical distribution $\hat{F}_{Y,T,X}$ in the same manner as $\hat{\theta}(t)$.

Theorem 6. *Under Assumptions 1 and A.1,*

$$\sqrt{n}(\hat{\theta}^*(t) - \hat{\theta}(t)) \rightsquigarrow_* \mathbb{V}_{\theta T}(t)$$

where \rightsquigarrow_* indicates weak convergence under the bootstrap law and $\mathbb{V}_{\theta T}$ is the tight mean 0 Gaussian process for each parameter $\theta(t)$ given above.

We can use the results of Theorem 6, to construct uniformly valid confidence bands that cover the entire curve with $(1 - \alpha)$ probability for any parameter of interest given by

$$\hat{C}_\theta(t) = \hat{\theta}(t) \pm \hat{c}_{1-\alpha} \hat{\Sigma}(t)^{1/2} / \sqrt{n}$$

where $\hat{c}_{1-\alpha}$ is a critical value satisfying

$$\lim_{n \rightarrow \infty} P(\theta(t) \in \hat{C}_\theta(t) \text{ for all } t \in \mathcal{T}) = 1 - \alpha$$

Here, $\hat{\Sigma}(t)$ denotes a uniformly consistent estimator of $\Sigma(t)$, the asymptotic variance function of $\sqrt{n}(\hat{\theta}(t) - \theta(t))$, such as

$$\hat{\Sigma}(t) = \frac{q_{0.75}(t) - q_{0.25}(t)}{z_{0.75}(t) - z_{0.25}(t)} \quad (4.1)$$

which is the bootstrap interquartile range scaled by the interquartile range of the standard normal distribution (this is a uniformly consistent estimate of $\Sigma(t)$, see Chernozhukov and Fernández-Val (2005)).

Consider the following bootstrap procedure. For some large number B and for each $b = 1, \dots, B$ compute

$$\hat{c}_b = \sup_{t \in \mathcal{T}} \hat{\Sigma}(t)^{-1/2} |\sqrt{n}(\hat{\theta}^b(t) - \hat{\theta}(t))|$$

where $\hat{\theta}^b(t)$ is the bootstrapped estimate of $\theta(t)$ using the b -th bootstrapped sample. Then, setting $\hat{c}_{1-\alpha}$ to be the $(1 - \alpha)$ quantile of $\{\hat{c}_b : 1 \leq b \leq B\}$ implies that $\hat{C}_\theta(t)$ asymptotically covers $\theta(t)$ for all values $t \in \mathcal{T}$ with probability $(1 - \alpha)$.

4.2 Testing if parameters depend on parents' income

We are also interested in testing whether each of the parameters of interest depends on t as discussed in Section 2.3. As in the previous section, let $\theta(t)$ generically denote one of the parameters of interest. The results in Theorems 2 to 5 imply that $\sqrt{n}(\hat{\theta}(t) - \theta(t)) \rightsquigarrow \mathbb{V}_{\theta T}$ in the space $l^\infty(\mathcal{T})$ where $\hat{\theta}(t)$ is the estimator of $\theta(t)$ and $\mathbb{V}_{\theta T}$ is some tight mean 0 Gaussian process that depends on which parameter is being estimated. Each of the parameters considered in the paper satisfies the following condition.

Condition 1. Denote any of the parameters considered above by $\theta(t)$ and its estimator given in the parameter by $\hat{\theta}(t)$. Also, for $\mu_\theta = E[\theta(T)]$, let

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_{\theta i} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta(T_i) - \mu_\theta)$$

denote the influence function for estimating μ_θ when θ is known. Then,

$$\left(\sqrt{n}(\hat{\theta}(t) - \theta(t)), \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_{\theta i} \right) \rightsquigarrow (\mathbb{V}_{\theta T}(t), \mathbb{W}_\theta)$$

in the space $l^\infty(\mathcal{T})$.

Consider $R_\theta(t) = \theta(t) - E[\theta(T)]$. In this section, we are interested in forming uniform confidence

bands for $R_\theta(t)$ as well as testing the null hypothesis

$$H_0 : R_\theta(t) = 0 \quad \text{for all } t \in \mathcal{T} \quad (4.2)$$

A natural estimator of $R_\theta(t)$ is given by $\hat{R}_\theta(t) = \hat{\theta}(t) - \frac{1}{n} \sum_{i=1}^n \hat{\theta}(T_i)$. The next result establishes the limiting process for $\hat{R}_\theta(t)$.

Proposition 1. *Under Assumptions 1 and A.1 and Condition 1*

$$\sqrt{n} \left(\hat{R}_\theta(t) - R_\theta(t) \right) \rightsquigarrow V_{\theta T}^R(t)$$

in the space $l^\infty(\mathcal{T})$ where $\mathbb{V}_{\theta T}^R(t)$ is a tight mean 0 Gaussian process given by $\mathbb{V}_{\theta T}^R(t) = \mathbb{V}_{\theta T}(t) + \int_{\mathcal{T}} \mathbb{V}_{\theta T}(t) \, dF_T(t) + \mathbb{W}_\theta$.

Proposition 1 can be used as the basis for constructing uniform confidence bands that asymptotically cover the entire curve with $(1 - \alpha)$ probability. To do this, we use the empirical bootstrap. Let $\hat{R}_\theta^*(t)$ denote the bootstrap version of $R_\theta(t)$. Given the result in Proposition 1, the following result follows

Corollary 1. *Under Assumptions 1 and A.1 and under Condition 1*

$$\sqrt{n} \left(\hat{R}_\theta^*(t) - \hat{R}_\theta(t) \right) \rightsquigarrow_* V_{\theta T}^R(t)$$

where $V_{\theta T}^R(t)$ is the Gaussian process given in Proposition 1.

The next corollary shows how to test H_0 given in Equation 4.2

Corollary 2. *Let $KS_\theta = \sup_{t \in \mathcal{T}} \Sigma_R(t)^{-1/2} |R_\theta(t)|$ and $\hat{KS}_\theta = \sup_{t \in \mathcal{T}} \hat{\Sigma}_R(t)^{-1/2} |\hat{R}_\theta(t)|$. Here, $\Sigma_R(t)$ is the asymptotic variance function of $\sqrt{n}(\hat{R}_\theta(t) - R_\theta(t))$ and $\hat{\Sigma}_R(t)$ is a uniformly consistent estimate of $\Sigma_R(t)$. Then, under H_0 (and under Assumptions 1 and A.1 and Condition 1),*

$$\sqrt{n} \left(\hat{KS}_\theta - KS_\theta \right) \rightsquigarrow \sup_{t \in \mathcal{T}} \Sigma_R(t)^{-1/2} |\mathbb{V}_{\theta T}^R(t)|$$

Moreover, let \hat{KS}_θ^* denote the bootstrapped version of KS_θ . Then,

$$\sqrt{n} \left(\hat{KS}_\theta^* - \hat{KS}_\theta \right) \rightsquigarrow_* \sup_{t \in \mathcal{T}} \Sigma_R(t)^{-1/2} |\mathbb{V}_{\theta T}^R|$$

Corollary 2 follows immediately from the continuous mapping theorem. It shows that one can test H_0 by comparing \hat{KS}_θ to a critical value given by the $(1 - \alpha)$ quantile of the bootstrapped $\sqrt{n} \left(\hat{KS}_\theta^* - \hat{KS}_\theta \right)$ which can be simulated a large number of times.

The last corollary of this section shows how to construct uniformly valid confidence bands for $R_\theta(t)$.

Corollary 3. *Under Assumptions 1 and A.1 and Condition 1 and consider the confidence region given by*

$$\hat{C}_\theta^R(t) = \hat{R}_\theta(t) \pm \hat{c}_{1-\alpha}^R \hat{\Sigma}_R(t)^{1/2} / \sqrt{n}$$

where $\hat{c}_{1-\alpha}^R$ is the $(1 - \alpha)$ quantile of $\sqrt{n} \left(\hat{R}_\theta^*(t) - \hat{R}_\theta(t) \right)$ which can be simulated a large number of times and where $\hat{\Sigma}_R(t)$ is the same as in Corollary 2. Then,

$$\lim_{n \rightarrow \infty} P(R_\theta(t) \in \hat{C}_R^\theta(t) \text{ for all } t \in \mathcal{T}) = 1 - \alpha$$

5 Data and results

5.1 Data description

The data that we use comes from the Panel Study of Income Dynamics (PSID) which has been the primary database used in much of the literature on intergenerational mobility. Like the majority of the income mobility literature using the PSID, we use total family income (including both father's and mother's income) instead of individual income (Chadwick and Solon 2002; Mayer and Lopoo 2005; Bloome 2015). The main alternative is to use only father's and son's income, but our approach offers several advantages. First, it seems likely that it is total family income that would affect a child's outcomes. Second, this approach allows us to keep daughters in the

analysis; in particular, families with one spouse with high income and the other with low income (or out of the labor force) will be treated as high income families in our analysis rather than as low income families.

The other main data issue in the intergenerational mobility literature is constructing measures of permanent income. Here, we follow existing work and use averages of income over several years to construct the permanent income (Solon 1992; Zimmerman 1992; Mazumder 2005). We construct child's permanent income (our outcome variable) in their adulthood by averaging at least three family incomes conditional on being at least 25 years old and being the head or the spouse of a household. We measure the parents' family income (our treatment variable) by averaging at least three family incomes when the child is 16 years old or younger. Before we calculate these family incomes, we drop yearly family incomes less than \$100. We also change all family incomes in all years into 2010 dollars using the CPI-U-RS series. More specifically, we first choose individuals whose ages are at least 1 in 1987 such that these individuals are at least 25 years old in 2011. Also, these individuals have to be less than 16 years old in 1970 to ensure that these individuals are sons or daughters at the very beginning of the survey. Finally, we drop the Survey of Economic Opportunity (SEO) part of the PSID sample; this is standard in the intergenerational mobility literature.

The covariates that we use in our analysis include child's gender and year of birth and the family head's gender, race, educational attainment, and veteran status. The main complication in obtaining the covariates of the family head is determining who is the family head, because the family head can change over time – for example, parents may divorce, remarry, or die over the course of their child's childhood. We set the family head characteristics as the mode of characteristics for the individual coded as the family head between the time that a child is born and reaches 16 years old. Our sample consists of 3,727 child-parent permanent family income pairs.

Table 1 provides summary statistics by quartile of parents' income. The 25th percentile of parents' income is \$44,200, the median is \$59,200, and the 75th percentile is \$78,200. As expected, child's income is increasing in parents' income. On average, children from families in the 1st, 2nd,

and 3rd quartiles have higher income than their parents; children from the fourth quartile have lower income on average than their parents.

There are some striking patterns in the data that are immediately noticeable, and most of these differences are most pronounced between the 1st and 2nd quartiles of parents' income. Parents in the first quartile are much more likely to be non-white than parents in the 2nd quartile (28% vs. 7%). Children from families in the 1st quartile are much less likely to have a male head (77% vs. 95%) which likely indicates that these children are from a single parent family.

Finally, there are big differences across parents' income quartiles in education. 36% of family heads in the lowest quartile have less than a high school education. The corresponding quantities are 20%, 9%, and 5% for the 2nd, 3rd, and 4th quartiles, respectively. There are also big differences in the fraction of heads with at least a college degree – 6% in the first quartile, 17% in the 2nd, 33% in the 3rd, and 57% in the 4th.

Taken together, the summary statistics suggest that child's income is positively correlated with parents' income. But child's income is also correlated with other background family characteristics – primarily education, race, and coming from a two-parent family – that are likely to also be important contributors to a child's income.

Table 2 presents OLS regression results of the log of child's income on the log of parents' income as well as additional controls. These results are useful to compare with the existing literature as well as to serve as a prelude to our main results. Without additional controls, the estimated IGE is 0.603.⁶ Adding demographic controls, as in specification (2) in the table, shrinks the estimated coefficient to 0.562. By far, the most important demographic control is a dummy variable for whether or not the race of a family is non-white. The third specification adds a control for year born which is likely to be important for the reasons mentioned earlier in this section; it has the expected sign but the estimate of the IGE does not change much. The fourth column adds education controls. Once again, the estimated IGE shrinks considerably to 0.443; so, here, adding additional controls reduced the estimated IGE by about 30%. The coefficients

⁶This estimate is towards the upper end of the range estimates of the IGE in the literature (Mazumder 2005; Black and Devereux 2011; Chetty, Hendren, Kline, and Saez 2014). However, recent work suggests that the IGE is larger using more recent periods, like in the current paper, than in earlier periods (Chetty et al. 2014; Davis and Mazumder 2017).

on the family head having less than a high school education and on the family head having at least a college degree (having a high school degree but less than a college degree is the omitted group) are large in magnitude. These results suggest that controlling for covariates such as race and education mitigates the effect of parents' income on child's income, though parents' income is still an important determinant of child's income.

5.2 Main Results

Our main results are provided in Figures 1 to 6 below. Each one corresponds to one of our main parameters of interest, and they each follow the same pattern. The top left panel provides the parameter as a function of parents' income from the observed data. The top right panel provides the same parameter as a function of parents' income but using the counterfactual distribution which adjusts for differences in observed covariates across different levels of parents' income. The bottom left panel shows the difference between the parameter coming from the observed distribution and the one coming from the counterfactual distribution. And the bottom right panel tests whether the parameter coming from the counterfactual distribution is the same across all values of parents' income. Each panel provides uniform confidence bands for the parameter of interest. This allows us to reject any hypothesis of interest for the entire function if the band does not cover 0.

Question 1: Does adjusting for covariates affect existing results?

The first part of our analysis considers very similar research questions as much existing work. We first focus on average child's income as a function of parents' income and its derivative which is a local version of the Intergenerational Elasticity (IGE) measure commonly reported in the intergenerational mobility literature. One difference between our results and most existing work is that our measure of the IGE is local, though Landersø and Heckman (2017) has considered a local IGE in previous work. However, the main departure in this section from existing work is that we also consider average child's income and local IGE *after adjusting for differences in covariates* across different levels of parents' income.

Average child's income as a function of parents' income and our measure of local IGE are reported in Figures 1 and 2. As expected, child's income is increasing in parents' income. This result holds using the observed distribution (top left panel) or after adjusting for differences in covariates (top right panel). On average, children from families with lower income have higher income than their parents while children from higher income families tend to have lower incomes than their parents. They cross at \$58,000 without adjusting for covariates and \$57,500 after adjusting for covariates.

Most interestingly, however, is that we can reject that adjusting for covariates does not make a difference in the estimates. Adjusting for covariates tends to increase expected income of children from low income families and decrease expected income of children from high income families (See the bottom left panel of Figure 1). This is in line with the results from the previous section where we saw that parents' income was strongly correlated with parents' education, parents' race, and having a male household head. It suggests that adjusting for differences in covariates decreases the strength of the relationship between parents' income and child's income.

Our estimates of Local IGEs are in Figure 2. Our unadjusted estimates indicate that the IGE may depend on parents' income. The estimated IGE is high (indicating low mobility) for children from low income families, and it is relatively lower for children from high income families (see the top left panel); for children from families at the poverty line, the estimated LIGE is 0.65 while for children from families in the 90th percentile of income, the estimated LIGE is 0.54. The confidence bands are wide though and we do not reject that the unadjusted LIGE is flat (results not shown in figure). Once other characteristics are adjusted for, the estimated LIGE is flat across parents' income levels; we do not reject that it is constant (see the bottom right panel). This result is particularly important as it says that apparent differences in mobility for children from low income families are likely explained by differences in other background characteristics. Second, the Local IGE is lower at all values of parents' income once one adjusts for difference in other characteristics. This result corresponds to the decrease in the estimated IGE as covariates are added in the regression results in Table 2.

Question 2: What is the effect of parents' income on the distribution of child's income? Average child's income conditional on parents' income only tells part of the story of the relationship. Our estimates in the previous section indicate that children from low income families have higher incomes than their parents on average. However, of course, not all children from families whose income takes a particular value have actual incomes equal to the average. Our methods allow us to look at these "distributional" parameters. First, we consider the effect of parents' income on the probability that a child's income is below the poverty line.

The results for the poverty rate are presented in Figure 3. Without adjusting for covariates, 24.3% children from families with incomes at the poverty line are estimated to have incomes below the poverty line themselves. After adjusting for covariates, only 15.7% are estimated to have incomes below the poverty line (all differences in this paragraph are statistically different from 0). At the median of parents' income, without adjusting for covariates 5.5% of children have income below the poverty line and slightly more, 5.9%, have income below the poverty line after adjusting for differences in observed characteristics. For children of families in the 90th percentile of income, we estimate that only 1.4% have incomes below the poverty line without adjusting for covariates while 2.4% have incomes below the poverty line when we do adjust for covariates. These results say that children from relatively poor families are much more likely to have incomes below the poverty line than children from middle or upper income families. This provides substantially more detail than simply looking at average child's income as a function of parents' income. In fact, children from relatively poor families do not just have lower incomes on average than children from other families, they are much more likely to have very low incomes themselves.

Similarly, children from low income families are much less likely to become "rich" than children from middle or high income families (we set the value to be considered "rich" at \$132,923 which is the 90th percentile of income in the U.S. in 2010). Without adjusting for covariates, we estimate that 1.0% of children from families at the poverty line, 6.0% of children from families at the median, and 21.5% of children from families at the 90th percentile become rich. Adjusting for covariates does not make much difference except for children from families at the 90th percentile where the estimate is reduced to 15.1%.

Next, we consider how wide the distribution of child’s income is as a function of parents’ income. To do this, we examine the variance of child’s income and the inter-quantile range of child’s income. First, Figure 5 plots the variance of child’s income as function of parents’ income. There are clear differences between the variance depending on whether or not the model adjusts for covariates. Without covariates, the variance of child’s income is higher for children with low income parents relative to high income parents (see the top left panel). However, once one accounts for differences in covariates across parents’ incomes, the variance flattens (see the top right panel and bottom right panel). Our results for the variance, however, are imprecise and we cannot reject that adjusting for covariates has no effect nor can we reject that the results that adjust for covariates do not change across parents’ incomes. On the other hand, we can reject that the variance is constant in the case where we do not adjust for covariates (results not shown in figure).

The inter-quantile range tells a similar story. These results are presented in Figure 6 (in the figure, we set $\tau_1 = 0.9$ and $\tau_2 = 0.1$). Without covariates, it appears that the spread of child’s income, as measured by the IQR, is decreasing in parents’ income. But adjusting for covariates instead indicates that the IQR is flat across parents’ incomes and that the differences are driven by differences for parents with very low income.

Summary of Main Results Our estimates of average child’s income as a function of parents’ income and of the Local IGE are largely in line with the existing literature. Children from families with relatively low income have lower earnings than children from higher income families. This result holds, though is somewhat reduced, when differences in covariates such as race and education are accounted for.

More interestingly, we were able to estimate the entire distribution of child’s income as a function of parent’s income. We found that children from families with low incomes were much more likely to have incomes below the poverty line than children from higher income families; again, this was somewhat mitigated when adjustments were made for differences in background characteristics, but there were still substantial differences. We also found suggestive evidence that the variance of child’s income was larger for children from low income families than from high income families, but adjusting for differences in covariates completely flattened the variance

across parents' income levels.

Relationship to Causal Estimates While we have argued that our estimates should not be considered to be estimates of the causal effect of parents' income on child's income, it is worth considering the relationship of our estimates to causal effect estimates. In the continuous treatment effect literature, it is common to impose the assumption of unconfoundedness; that is,

$$Y(t) \perp\!\!\!\perp T|X \quad \forall t \in \mathcal{T}$$

where $Y(t)$ denotes an individual's "potential outcome" if that individual received treatment t . In this case, the distribution dose-response function is $F_{Y|T}^C$ – exactly the same as our counterfactual distribution. And all of our parameters of interest are dose-response functions as well. Then, for example, the average treatment effect is given by

$$ATE(t, t') = E^C[Y|T = t] - E^C[Y|T = t']$$

and depends on two values of parents' income. This is a parameter that we have only talked about indirectly in our main analysis. Set $t = t_{0.75}$ and $t' = t_{0.25}$ which represent the 75th percentile and 25th percentile of parents' income, respectively. $ATE(t_{0.75}, t_{0.25})$ is how much a random individual's income would increase on average if they changed from having parents in the 25th percentile of the income distribution to the 75th percentile. We estimate that, on average, moving from the 25th percentile to the 75th percentile of parents' income increases child's income by 23.5 log points. Under the assumption of unconfoundedness, this should be interpreted as a causal effect. Similarly, we estimate that, under unconfoundedness, moving from the 25th percentile to the 75th percentile of family income decreases the probability of a child's income being below the poverty line by 3.60 percentage points (a 45% reduction).

We suspect that, in the context of intergenerational income mobility, estimates of the causal effect using our approach are likely to overstate the effect of parents' income on child's income. This would be the case if children of high income parents have some latent characteristics (or their

parents have some latent characteristics) that lead to higher income relative to children of low income parents even after conditioning on observables. One small piece of evidence related to this concerns parents' education. For education, we include three dummy variables – less than high school, high school graduate but not a college graduate, or a college graduate. Looking within these groups, parents in the top quartile have more education than parents from the bottom quartile; for example, parents with a college degree from the top quartile are relatively more likely to have an advanced degree and parents with a high school degree are relatively more likely to have some college than parents in the bottom quartile. Likewise, we suspect that our estimates of the effect of parents' income on the probability of a child having income below the poverty line will overstate the causal effect of parents' income for similar reasons. It is less clear the direction of the bias for estimating the spread parameters, such as the variance of child's incomes, conditional on parents' income.

6 Conclusion

This paper has developed new tools to study intergenerational income mobility. Our methods allow us to (1) study the entire distribution of child's income conditional on parents' income, (2) adjust for differences in observed characteristics among children who have parents with different income levels, and (3) treat parents' income as a continuous variable rather than splitting it into a small number of groups. These tools may be useful to researchers in other fields who are interested in counterfactual distributions with a continuous treatment or are interested in the causal effect of a continuous treatment under the assumption of unconfoundedness.

In line with the existing literature on intergenerational income mobility, we found that average child's income is increasing in parents' income and that children from low income families have higher incomes than their parents on average. Using new tools developed in the paper, we were able to learn considerably more about the effect of parents' income on children's income by identifying and estimating the distribution of child's income as a function of parents' income. We found that the fraction of children with income below the poverty line declined sharply in parents' income. We documented that the distribution of observed characteristics is quite different across

the distribution of parents' income. Adjusting for these differences decreased the correlation between child's income and parents' income and decreased the effect of parents' income on the fraction of children that are in poverty for children from low income families while increasing it for high income families. None of these results would have been available using standard tools to analyze intergenerational income mobility.

References

- [1] Bhattacharya, Debopam and Bhashkar Mazumder. “A nonparametric analysis of black–white differences in intergenerational income mobility in the United States”. *Quantitative Economics* 2.3 (2011), pp. 335–379.
- [2] Björklund, Anders, Mikael Lindahl, and Erik Plug. “The origins of intergenerational associations: Lessons from Swedish adoption data”. *The Quarterly Journal of Economics* 121.3 (2006), pp. 999–1028.
- [3] Björklund, Anders, Jesper Roine, and Daniel Waldenström. “Intergenerational top income mobility in Sweden: Capitalist dynasties in the land of equal opportunity?” *Journal of Public Economics* 96.5 (2012), pp. 474–484.
- [4] Black, Sandra E and Paul Devereux. “Recent Developments in Intergenerational Mobility, Handbook of Labor Economics, Amsterdam, North-Holland”. *Handbook of Labor Economics*. Ed. by Ashenfelter, Orley and David Card. Vol. 4.B. Amsterdam: North-Holland, 2011. Chap. 16, pp. 1487–1541.
- [5] Blanden, Jo, Paul Gregg, and Lindsey Macmillan. “Accounting for intergenerational persistence”. *Economic Journal* 117 (2007), pp. C43–C60.
- [6] Bloome, Deirdre. “Income inequality and intergenerational income mobility in the United States”. *Social Forces* 93.3 (2015), pp. 1047–1080.
- [7] Bowles, Samuel and Herbert Gintis. “The inheritance of inequality”. *The Journal of Economic Perspectives* 16.3 (2002), pp. 3–30.
- [8] Bratberg, Espen, Øivind Anti Nilsen, and Kjell Vaage. “Job losses and child outcomes”. *Labour Economics* 15.4 (2008), pp. 591–603.
- [9] Bratsberg, Bernt, Knut Røed, Oddbjørn Raaum, Robin Naylor, Tor Eriksson, et al. “Nonlinearities in intergenerational earnings mobility: consequences for cross-country comparisons”. *The Economic Journal* 117.519 (2007).
- [10] Chadwick, Laura and Gary Solon. “Intergenerational income mobility among daughters”. *The American Economic Review* 92.1 (2002), pp. 335–344.
- [11] Chernozhukov, Victor and Iván Fernández-Val. “Subsampling inference on quantile regression processes”. *Sankhyā: The Indian Journal of Statistics* (2005), pp. 253–276.
- [12] Chernozhukov, Victor, Iván Fernández-Val, and Alfred Galichon. “Quantile and probability curves without crossing”. *Econometrica* 78.3 (2010), pp. 1093–1125.
- [13] Chernozhukov, Victor, Ivan Fernandez-Val, and Blaise Melly. “Inference on counterfactual distributions”. *Econometrica* 81.6 (2013), pp. 2205–2268.

- [14] Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. “The fading American dream: Trends in absolute income mobility since 1940”. *Science* 356.6336 (2017), pp. 398–406.
- [15] Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. “Where is the land of opportunity? The geography of intergenerational mobility in the United States”. *The Quarterly Journal of Economics* 129.4 (2014), pp. 1553–1623.
- [16] Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. “Is the United States still a land of opportunity? Recent trends in intergenerational mobility”. *The American Economic Review* 104.5 (2014), pp. 141–147.
- [17] Collins, William J and Marianne H Wanamaker. “Up from Slavery? African American Intergenerational Economic Mobility Since 1880”. Working Paper. 2017.
- [18] Currie, Janet and Douglas Almond. “Human capital development before age five”. *Handbook of labor economics* 4 (2011), pp. 1315–1486.
- [19] Dahl, Gordon B and Lance Lochner. “The impact of family income on child achievement: Evidence from the earned income tax credit”. *The American Economic Review* 102.5 (2012), pp. 1927–1956.
- [20] Davis, Jonathan and Bhashkar Mazumder. “The Decline in Intergenerational Mobility After 1980”. Working Paper. 2017.
- [21] DiNardo, John, Nicole M Fortin, and Thomas Lemieux. “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach”. *Econometrica* 64.5 (1996), pp. 1001–1044.
- [22] Eide, Eric R and Mark H Showalter. “Factors affecting the transmission of earnings across generations: A quantile regression approach”. *Journal of Human Resources* (1999), pp. 253–267.
- [23] Firpo, Sergio. “Efficient Semiparametric Estimation of Quantile Treatment Effects”. *Econometrica* 75.1 (2007), pp. 259–276.
- [24] Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux. “Unconditional quantile regressions”. *Econometrica* 77.3 (2009), pp. 953–973.
- [25] Florens, Jean-Pierre, James J Heckman, Costas Meghir, and Edward Vytlacil. “Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects”. *Econometrica* 76.5 (2008), pp. 1191–1206.
- [26] Flores, Carlos A. “Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment”. Working Paper. 2007.

- [27] Foresi, Silverio and Franco Peracchi. “The conditional distribution of excess returns: An empirical analysis”. *Journal of the American Statistical Association* 90.430 (1995), pp. 451–466.
- [28] Galvao, Antonio F and Liang Wang. “Uniformly Semiparametric Efficient Estimation of Treatment Effects With a Continuous Treatment”. *Journal of the American Statistical Association* 110.512 (2015), pp. 1528–1542.
- [29] Grawe, Nathan D. “Reconsidering the use of nonlinearities in intergenerational earnings mobility as a test for credit constraints”. *Journal of Human Resources* 39.3 (2004), pp. 813–827.
- [30] Groves, Melissa Osborne. “Personality and the intergenerational transmission of economic status”. *Unequal chances: Family background and economic success* (2005), pp. 208–231.
- [31] Hirano, Keisuke and Guido W Imbens. “The propensity score with continuous treatments”. *Applied Bayesian modeling and causal inference from incomplete-data perspectives* 226164 (2004), pp. 73–84.
- [32] Jantti, Markus et al. “American exceptionalism in a new light: A comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States”. Working Paper. 2006.
- [33] Kennedy, Edward H, Zongming Ma, Matthew D McHugh, and Dylan S Small. “Non-parametric methods for doubly robust estimation of continuous treatment effects”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2016).
- [34] Koenker, Roger. *Quantile Regression*. Cambridge University Press, 2005.
- [35] Koenker, Roger and Gilbert Bassett Jr. “Regression Quantiles”. *Econometrica* (1978), pp. 33–50.
- [36] Kosorok, Michael R. *Introduction to empirical processes and semiparametric inference*. Springer Science & Business Media, 2007.
- [37] Landersø, Rasmus and James J Heckman. “The scandinavian fantasy: The sources of intergenerational mobility in denmark and the U.S.” *The Scandinavian Journal of Economics* 119.1 (2017), pp. 178–230.
- [38] Liu, Haoming and Jinli Zeng. “Genetic ability and intergenerational earnings mobility”. *Journal of Population Economics* 22.1 (2009), pp. 75–95.
- [39] Machado, José AF and José Mata. “Counterfactual decomposition of changes in wage distributions using quantile regression”. *Journal of applied Econometrics* 20.4 (2005), pp. 445–465.
- [40] Mayer, Susan E and Leonard M Lopoo. “Has the intergenerational transmission of economic status changed?” *Journal of Human Resources* 40.1 (2005), pp. 169–185.

- [41] Mazumder, Bhashkar. “Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data”. *Review of Economics and Statistics* 87.2 (2005), pp. 235–255.
- [42] Melly, Blaise. “Decomposition of differences in distribution using quantile regression”. *Labour economics* 12.4 (2005), pp. 577–590.
- [43] Oreopoulos, Philip, Marianne Page, and Ann Huff Stevens. “The intergenerational effects of worker displacement”. *Journal of Labor Economics* 26.3 (2008), pp. 455–483.
- [44] Richey, Jeremiah and Alicia Rosburg. “Decomposing economic mobility transition matrices”. Working Paper. 2015.
- [45] Richey, Jeremiah and Alicia Rosburg. “Changing roles of ability and education in US intergenerational mobility”. *Economic Inquiry* 55.1 (2017), pp. 187–201.
- [46] Rothe, Christoph. “Nonparametric estimation of distributional policy effects”. *Journal of Econometrics* 155.1 (2010), pp. 56–70.
- [47] Solon, Gary. “Intergenerational income mobility in the United States”. *The American Economic Review* (1992), pp. 393–408.
- [48] Solon, Gary. “Intergenerational mobility in the labor market”. *Handbook of labor economics* 3 (1999), pp. 1761–1800.
- [49] Van Der Vaart, Aad W and Jon A Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, 1996.
- [50] Zimmerman, David J. “Regression toward mediocrity in economic stature”. *The American Economic Review* (1992), pp. 409–429.

A Additional Assumptions

Assumption A.1.

(i) \mathcal{YTX} , which denotes Cartesian product of the supports of Y , T , and X , is a compact subset of \mathbb{R}^{2+k} where k is the dimension of X .

(ii) Y is continuously distributed with conditional density $f_{Y|T,X}(y|t, x)$ uniformly bounded away from 0 and ∞ and continuous in $(y, t, x) \in \mathcal{YTX}$.

(iii) The support \mathcal{T} of T is the compact interval $[t_{\min}, t_{\max}]$ with density $f_T(t)$ bounded away from 0 and ∞ on \mathcal{T} .

(iv) For $\mathcal{U} \subset (0, 1)$, $F_{Y|T}$ and $F_{Y|T}^C$ admit positive continuous densities $f_{Y|T}$ and $f_{Y|T}^C$ on an interval $[a, b]$ containing an ϵ -enlargement of the sets $\{Q_{Y|T}(\tau|t) : \tau \in \mathcal{U}\}$ and $\{Q_{Y|T}^C(\tau|t) : \tau \in \mathcal{U}\}$, respectively.

(v) $E\| [TX']' \|^2 < \infty$

(vi) Let $J_1(y) = E \left[\frac{\lambda(\alpha_{11}(y)T + X'\alpha_{12}(y))^2}{\Lambda(\alpha(y)T + X'\beta(y))(1 - \Lambda(\alpha(y)T + X'\beta(y)))} [TX']' [TX'] \right]$. Also, let

$J_2(y) = E \left[\frac{\lambda(\alpha_{20} + \alpha_{21}(y)T)^2}{\Lambda(\alpha_{20} + \alpha_{21}(y)T)(1 - \Lambda(\alpha_{20}(y) + \alpha_{21}(y)T))} [1T]' [1T] \right]$. The minimum eigenvalues of $J_1(y)$ and $J_2(y)$ are uniformly bounded away from zero.

B Additional Asymptotic Results

This appendix contains additional limiting processes for (i) the variance of child's income conditional on parents' income and (ii) the inter-quantile range of child's income conditional on parents' income.

The first result establishes the limiting process for the variance of child's income conditional on family income.

Theorem B.1. *Under Assumptions 1 and A.1,*

$$\sqrt{n}(\hat{V}ar(Y|T=t) - Var(Y|T=t)) \rightsquigarrow \mathbb{V}_T^V(t)$$

in the space $l^\infty(\mathcal{T})$ where \mathbb{V}_T^V is tight Gaussian process with mean 0 that is given by

$$\mathbb{V}_T^V(t) = 2 \int_0^1 (Q_{Y|T}(\tau|t) - E[Y|T=t]) \left(\mathbb{Z}(\tau, t) - \int_0^1 \mathbb{Z}(u, t) \, du \right) \, d\tau$$

In addition,

$$\sqrt{n}(\hat{V}ar^C(Y|T=t) - Var^C(Y|T=t)) \rightsquigarrow \mathbb{V}_T^{C,V}(t)$$

in the space $l^\infty(\mathcal{T})$ where $\mathbb{V}_T^{C,V}$ is a tight Gaussian process with mean 0 that is given by

$$\mathbb{V}_T^{C,V}(t) = 2 \int_0^1 (Q_{Y|T}^C(\tau|t) - E^C[Y|T=t]) \left(\mathbb{Z}^C(\tau, t) - \int_0^1 \mathbb{Z}^C(u, t) \, du \right) \, d\tau$$

Finally,

$$\sqrt{n}(\hat{\Delta}^{Var}(t) - \Delta^{Var}(t)) \rightsquigarrow \mathbb{V}_T^{\Delta, Var}(t)$$

where $\mathbb{V}_T^{\Delta, Var}$ is a tight Gaussian process with mean 0 that is given by $\mathbb{V}_T^{\Delta, Var} = \mathbb{V}_T^V - \mathbb{V}_T^{C,V}$.

The final theorem in this section provides the limiting process of the inter-quantile range of child's income conditional on family income.

Theorem B.2. *Under Assumptions 1 and A.1,*

$$\sqrt{n}(\hat{IQR}(\tau_1, \tau_2, t) - IQR(\tau_1, \tau_2, t)) \rightsquigarrow \mathbb{G}_T^{IQR}(t)$$

in the space $l^\infty(\mathcal{T})$ where \mathbb{G}_T^{IQR} is a tight mean 0 Gaussian process given by $\mathbb{G}_T^{IQR}(\tau_1, \tau_2, t) = \mathbb{Z}(\tau_1|t) - \mathbb{Z}(\tau_2|t)$ where \mathbb{Z} is given in Theorem 3. Also,

$$\sqrt{n}(\hat{IQR}^C(\tau_1, \tau_2, t) - IQR^C(\tau_1, \tau_2, t)) \rightsquigarrow \mathbb{G}_T^{C, IQR}(t)$$

in the space $l^\infty(\mathcal{T})$ where $\mathbb{G}_T^{C, IQR}$ is a tight Gaussian process with mean 0 given by $\mathbb{G}_T^{C, IQR}(\tau_1, \tau_2, t) = \mathbb{Z}^C(\tau_1|t) - \mathbb{Z}^C(\tau_2|t)$ where \mathbb{Z}^C is given in Theorem 3. Finally,

$$\sqrt{n}(\hat{\Delta}^{IQR}(\tau_1, \tau_2, t) - \Delta^{IQR}(\tau_1, \tau_2, t)) \rightsquigarrow \mathbb{G}_T^{IQR}(\tau_1, \tau_2, t) - \mathbb{G}_T^{C, IQR}(\tau_1, \tau_2, t)$$

C Proofs

C.1 Proof of Theorem 1

Let

$$\hat{G}_{Y|T,X}(y|t, x) = \sqrt{n}(\hat{F}_{Y|T,X}(y|t, x) - F_{Y|T,X}(y|t, x))$$

and let

$$\hat{G}_X(x) = \sqrt{n}(\hat{F}_X(x) - F_X(x))$$

which are the empirical processes of the conditional distribution of child's income and the distribution of other observable characteristics.

Also, let $\xi = (y, t, x, \bar{y}, \bar{t}, \bar{x})$ and $W = (Y, T, X)$ and let

$$\psi_1(W, \xi) = -\lambda(\alpha_{20}(y) + \alpha_{21}(y)t) \begin{bmatrix} 1 & t \end{bmatrix} J_1(\bar{y})^{-1} H_1(Y, T, y)$$

where λ is the derivative of the link function Λ and

$$H_1(Y, T, y) = (\mathbb{1}\{Y \leq y\} - \Lambda(\alpha_{20}(y) + \alpha_{21}(y)T)) \times \frac{\lambda(\alpha_{20}(y) + \alpha_{10}(y)T)}{(\Lambda(\alpha_{20}(y) + \alpha_{10}(y)T))(1 - \Lambda(\alpha_{20}(y) + \alpha_{10}(y)T))} \begin{bmatrix} 1 \\ T \end{bmatrix}$$

Next, let

$$\psi_2(W, \xi) = -\lambda(\alpha_{10}(\bar{y}) + \alpha_{12}(\bar{y})\bar{t} + \bar{x}'\alpha_{12}(\bar{y})) \begin{bmatrix} 1 & \bar{t} & \bar{x}' \end{bmatrix} J(\bar{y})^{-1} H_2(Y, T, X, \bar{y})$$

where

$$H_2(Y, T, X, y) = (\mathbb{1}\{Y \leq y\} - \Lambda(\alpha_{10}(y) + \alpha_{11}(y)T + X'\alpha_{12}(y))) \times \frac{\lambda(\alpha_{10}(y) + \alpha_{11}(y)T + X'\alpha_{12}(y))}{(\Lambda(\alpha_{10}(y) + \alpha_{11}(y)T + X'\alpha_{12}(y)))(1 - \Lambda(\alpha_{10}(y) + \alpha_{11}(y)T + X'\alpha_{12}(y)))} \begin{bmatrix} 1 \\ T \\ X \end{bmatrix}$$

and let

$$\psi_3(W, \xi) = \mathbb{1}\{X \leq x\} - F_X(x)$$

The first result establishes the joint limiting distribution of $\hat{G}_{Y|T}$, $\hat{G}_{Y|T,X}$, and \hat{G}_X .

Lemma 1. *Let $\mathbb{S} = l^\infty(\mathcal{YT}) \times l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X})$. Under Assumption 1 and Assumption A.1,*

$$(\hat{G}_{Y|T}(y|t), \hat{G}_{Y|X,T}(\bar{y}|\bar{x}, \bar{t}), \hat{G}_X(x)) \rightsquigarrow (\mathbb{G}_{Y|T}, \mathbb{G}_{Y|T,X}, \mathbb{G}_X)$$

in the space \mathbb{S} and where $(\mathbb{G}_{Y|T,X}, \mathbb{G}_X)$ is a tight Gaussian process with mean 0 with covariance function $V(\xi_1, \xi_2)$ defined on \mathbb{S} and given by

$$V(\xi_1, \xi_2) = E[\psi(W, \xi_1)\psi(W, \xi_2)']$$

where $\psi(W, \xi) = (\psi_1(W, \xi), \psi_2(W, \xi), \psi_3(W, \xi))'$

Proof. The result follows immediately under Assumptions 1 and A.1 and from the results in Chernozhukov, Fernandez-Val, and Melly (2013). \square

Before proving the main result, we consider the following result first.

Lemma 2. Consider the map $\psi : \mathbb{D}_\psi \subset \mathbb{D} = l^\infty(\mathcal{Y}\mathcal{T}\mathcal{X}) \times l^\infty(\mathcal{X}) \mapsto l^\infty(\mathcal{Y}\mathcal{T})$ given by

$$\psi(\Lambda) = \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x) \, d\Lambda_2(x)$$

for $\Lambda = (\Lambda_1, \Lambda_2) \in \mathbb{D}$. Then, the map ψ is Hadamard differentiable at Λ_0 tangentially to \mathbb{D} with derivative at Λ_0 in $\lambda = (\lambda_1, \lambda_2) \in \mathbb{D}$ given by

$$\psi'_{\Lambda_0}(\lambda) = \int_{\mathcal{X}} \lambda_1(\cdot|\cdot, x) \, d\Lambda_{20}(x) + \int_{\mathcal{X}} \Lambda_{10}(\cdot|\cdot, x) \, d\lambda_2(x)$$

Proof. Consider any sequence $t_k > 0$ and $\Lambda_k \in \mathbb{D}$ for $k = 1, 2, 3, \dots$ with $t_k \downarrow 0$ and

$$\begin{aligned} \lambda_{1k} &= \frac{\Lambda_{1k} - \Lambda_1}{t_k} \\ \lambda_{2k} &= \frac{\Lambda_{2k} - \Lambda_2}{t_k} \end{aligned}$$

with $(\lambda_{1k}, \lambda_{2k}) \rightarrow (\lambda_1, \lambda_2) \in \mathbb{D}$ as $k \rightarrow \infty$.

Then,

$$\begin{aligned} \frac{\psi(\Lambda_k) - \psi(\Lambda)}{t_k} - \psi'_\Lambda(\lambda) &= \int_{\mathcal{X}} \Lambda_{1k}(\cdot|\cdot, x) \, d\Lambda_{2k}(x)/t_k - \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x) \, d\Lambda_2(x)/t_k \\ &\quad - \int_{\mathcal{X}} \lambda_1(\cdot|\cdot, x) \, d\Lambda_2(x) - \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x) \, d\lambda_2(x) \\ &= \int_{\mathcal{X}} \frac{\Lambda_{1k}(\cdot|\cdot, x) - \Lambda_1(\cdot|\cdot, x)}{t_k} \, d(\Lambda_{2k}(x) - \Lambda_2(x)) \\ &\quad + \int_{\mathcal{X}} \frac{\Lambda_{1k}(\cdot|\cdot, x) - \Lambda_1(\cdot|\cdot, x)}{t_k} \, d\Lambda_2(x) \\ &\quad + \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x) \, d(\Lambda_{2k}(x) - \Lambda_2(x))/t_k \\ &\quad - \int_{\mathcal{X}} \lambda_1(\cdot|\cdot, x) \, d\Lambda_2(x) - \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x) \, d\lambda_2(x) \\ &= t_k \int_{\mathcal{X}} \lambda_{1k}(\cdot|\cdot, x) \, d\lambda_{2k}(x) \\ &\quad + \int_{\mathcal{X}} (\lambda_{1k}(\cdot|\cdot, x) - \lambda_1(\cdot|\cdot, x)) \, d\Lambda_2(x) \\ &\quad + \int_{\mathcal{X}} \Lambda_1(\cdot|\cdot, x) \, d(\lambda_{2k} - \lambda_2)(x) \\ &\rightarrow 0 \text{ as } k \rightarrow \infty \end{aligned}$$

where, in the last equation, the first line is $O(t_k)$ which converges to 0 as $k \rightarrow \infty$, and the second and third terms converge to 0 because $(\lambda_{1k}, \lambda_{2k}) \rightarrow (\lambda_1, \lambda_2)$. \square

Lemma 3. Consider the map $\phi : \mathbb{D}_\phi \subset l^\infty(\mathcal{YT}) \times l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X}) \mapsto l^\infty(\mathcal{YT})^2$ given by

$$\phi(\Gamma) = (\Gamma_1, \psi(\Gamma_2, \Gamma_3))$$

in $\Gamma = (\Gamma_1, \Gamma_2, \Gamma_3) \in l^\infty(\mathcal{YT}) \times l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X})$ and the map $\psi : \mathbb{D}_\psi \subset l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X}) \mapsto l^\infty(\mathcal{YT})$ is given in Lemma 2. Then, the map ϕ is Hadamard differentiable at Γ_0 tangentially to $l^\infty(\mathcal{YT}) \times l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X})$ with derivative at Γ_0 in $\gamma = (\gamma_1, \gamma_2, \gamma_3) \in \mathbb{D}$ given by

$$\begin{aligned} \phi'_{\Gamma_0}(\gamma) &= (\gamma_1, \psi'_{(\Gamma_{20}, \Gamma_{30})}(\gamma_2, \gamma_3)) \\ &= \left(\gamma_1, \int_{\mathcal{X}} \gamma_2(\cdot | \cdot, x) \, d\Gamma_{30}(x) + \int_{\mathcal{X}} \Gamma_{20}(\cdot | \cdot, x) \, d\gamma_3(x) \right) \end{aligned}$$

Proof. The result follows immediately from Lemma 2. □

Proof of Theorem 1

Lemma 3 implies

$$(\hat{G}_{Y|T}(y|t), \hat{G}_{Y|T}^C(\bar{y}|\bar{t})) \rightsquigarrow (\mathbb{G}_{Y|T}, \mathbb{G}_{Y|T}^C)$$

indexed by (y, t, \bar{y}, \bar{t}) in $\mathbb{S} = l^\infty(\mathcal{YT})^2$ and where $\mathbb{G}_{Y|T}$ is given in Lemma 1 and

$$\mathbb{G}_{Y|T}^C = \int_{\mathcal{X}} \mathbb{G}_{Y|T,X}(\cdot | \cdot, x) \, dF_X(x) + \int_{\mathcal{X}} F_{Y|T,X}(\cdot | \cdot, x) \, d\mathbb{G}_X(x)$$

($\mathbb{G}_{Y|T,X}$ and \mathbb{G}_X are given in Lemma 1). Then, the process given in Theorem 1 is given by setting $\bar{y} = y$ and $\bar{t} = t$.

C.2 Proof of Theorem 2

The result follows immediately from Theorem 1.

C.3 Proof of Theorem 3

The result follows from Theorem 2 and by Lemma 3.9.23(ii) in Van Der Vaart and Wellner (1996).

C.4 Proof of Theorem 4

Lemma 4. Let $\mathbb{D} = l^\infty(\mathcal{UT})$ and define the map $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto l^\infty(\mathcal{T})$ given by

$$\phi(\Phi) = \int_0^1 \Phi(\tau | \cdot) d\tau$$

for $\Phi \in \mathbb{D}$. Then, the map ϕ is Hadamard differentiable at Φ_0 tangentially to \mathbb{D} with derivative at Φ_0 in $\xi \in \mathbb{D}$ given by

$$\phi'_{\Gamma_0}(\xi) = \int_0^1 \xi(\tau|\cdot) d\tau$$

Proof. Consider any sequence $t_k > 0$ and $\Phi_k \in \mathbb{D}$ for $k = 1, 2, 3, \dots$ with $t_k \downarrow 0$ and

$$\xi_k = \frac{\Phi_k - \Phi}{t_k} \rightarrow \xi \in \mathbb{D} \text{ as } k \rightarrow \infty$$

Then,

$$\begin{aligned} \frac{\phi(\Phi_k) - \phi(\Phi)}{t_k} - \phi'_{\Phi}(\xi) &= \int_0^1 \frac{\Phi_k(\tau|\cdot) - \Phi(\tau|\cdot)}{t_k} d\tau - \int_0^1 \xi(\tau|\cdot) d\tau \\ &= \int_0^1 \xi_k(\tau|\cdot) - \xi(\tau|\cdot) d\tau \\ &\leq \|\xi_k - \xi\|_{\infty} \int_0^1 d\tau \rightarrow 0 \text{ as } k \rightarrow \infty \end{aligned}$$

□

Proof of Theorem 4 For the first part,

$$\begin{aligned} \sqrt{n}(\hat{E}[Y|T=t] - E[Y|T=t]) &= \sqrt{n} \left(\frac{1}{S} \sum_{s=1}^S \hat{Q}_{Y|T}(\tau_s|t) - \phi(\hat{Q}_{Y|T}) \right) + \sqrt{n}(\phi(\hat{Q}_{Y|T}) - \phi(Q_{Y|T})) \\ &= \phi'_{Q_{Y|T}} \sqrt{n}(\hat{Q}_{Y|T} - Q_{Y|T}) + o_p(1) \end{aligned}$$

which holds uniformly in t under the condition that S is large enough, e.g. $S = Cn^{1/2+\delta}$ for some $C > 0$ and $\delta > 0$, and by Lemma 4. This implies the result.

The second and third parts follows exactly the same argument.

C.5 Proof of Proposition 1

Lemma 5. Consider the map $\psi : \mathbb{D}_{\psi} \subset l^{\infty}(\mathcal{T}) \mapsto \mathbb{R}$ given by

$$\psi(\Lambda) = \int_{\mathcal{T}} \Lambda(t) dF_T(t)$$

in $\Lambda \in l^{\infty}(\mathcal{T})$. Then, the map ψ is Hadamard differentiable at Λ_0 tangentially to $l^{\infty}(\mathcal{T})$ with derivative at Λ_0 in $\lambda \in l^{\infty}(\mathcal{T})$ given by

$$\psi'_{\Lambda_0}(\lambda) = \int_{\mathcal{T}} \lambda(t) dF_T(t)$$

Proof. The proof of this result follows using essentially the same arguments as in Lemma 4, though this case is somewhat easier. \square

Proof of Proposition 1

$$\begin{aligned}\sqrt{n} \left(\hat{R}_\theta(t) - R_\theta(t) \right) &= \sqrt{n} \left(\hat{\theta}(t) - \int_{\mathcal{T}} \hat{\theta}(t) d\hat{F}_T(t) \right) - \sqrt{n} \left(\theta(t) - \int_{\mathcal{T}} \theta(t) dF_T(t) \right) \\ &= \sqrt{n}(\hat{\theta}(t) - \theta(t))\end{aligned}\tag{C.1}$$

$$+ \sqrt{n} \int_{\mathcal{T}} \hat{\theta}(t) - \theta(t) d\hat{F}_T(t)\tag{C.2}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \theta(T_i) - E[\theta(T)]\tag{C.3}$$

The term in Equation C.1 weakly converges to $\mathbb{V}_{\theta T}$ and the term in Equation C.3 weakly converges to \mathbb{W} – both of these hold by Condition 1.

For Equation C.2, note that

$$\begin{aligned}\sqrt{n} \int_{\mathcal{T}} \hat{\theta}(t) - \theta(t) d\hat{F}_T(t) &= \sqrt{n} \int_{\mathcal{T}} \hat{\theta}(t) - \theta(t) d(\hat{F}_T - F_T)(t) + \sqrt{n} \int_{\mathcal{T}} \hat{\theta}(t) - \theta(t) dF_T(t) \\ &= \sqrt{n} \int_{\mathcal{T}} \hat{\theta}(t) - \theta(t) dF_T(t) + o_p(1)\end{aligned}$$

which weakly converges to $\int_{\mathcal{T}} V_{\theta T}(t) dF_T(t)$ by Lemma 5. This implies the result.

C.6 Proof of Theorem 6

As a first step, we prove the following lemma.

Lemma 6. *Under Assumptions 1 and A.1,*

$$(\hat{G}_{Y|T}^*(y|t), \hat{G}_{Y|T}^{C*}(y|t)) \rightsquigarrow_* (\mathbb{V}_{Y|T}, \mathbb{V}_{Y|T}^C)$$

where \rightsquigarrow_* indicates weak convergence under the bootstrap law and $(\mathbb{V}_{Y|T}, \mathbb{V}_{Y|T}^C)$ is the Gaussian process from Theorem 1.

Proof. The result follows from Theorem 1 and by Theorem 3.6.1 of Van Der Vaart and Wellner (1996). \square

Proof of Theorem 6: The result follows from Lemma 6, that $\theta(t)$ is a Hadamard differentiable function of the $F_{Y|T}$ and $F_{Y|T}^C$, and by the functional delta method applied to the bootstrap.

C.7 Proof of Theorem B.1

Lemma 7. Let $\mathbb{D} = l^\infty(\mathcal{UT})$ and define the map $\pi : \mathbb{D}_\pi \subset \mathbb{D} \mapsto l^\infty(\mathcal{T})$ by

$$\pi(\Gamma) = \int_0^1 \left(\Gamma(\tau|\cdot) - \int_0^1 \Gamma(u|\cdot) \, du \right)^2 \, d\tau$$

for $\Gamma \in \mathbb{D}$. Then, the map π is Hadamard differentiable at Γ_0 tangentially to \mathbb{D} with derivative at Γ_0 in $\gamma \in \mathbb{D}$ given by

$$\pi'_{\Gamma_0}(\gamma) = 2 \int_0^1 \left\{ \left(\Gamma_0(\tau|\cdot) - \int_0^1 \Gamma_0(u|\cdot) \, du \right) \left(\gamma(\tau|\cdot) - \int_0^1 \gamma(u|\cdot) \, du \right) \right\} \, d\tau$$

Proof. Consider the maps $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto l^\infty(\mathcal{T})$ given in Lemma 4, $\pi_1 : \mathbb{D}_{\pi_1} \subset l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T}) \mapsto \mathbb{D}_{\pi_2}$ given by

$$\pi_1(\Lambda) = \Lambda_1 - \Lambda_2$$

for $\Lambda = (\Lambda_1, \Lambda_2) \in l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T})$ and the map $\pi_2 : \mathbb{D}_{\pi_2} \subset l^\infty(\mathcal{UT}) \mapsto \mathbb{D}_\phi$ given by

$$\pi_2(\Theta) = \Theta^2$$

for $\Theta \in l^\infty(\mathcal{UT})$

First, notice that the map π is given by the composition map $\pi(\Gamma) = \phi \circ \pi_2 \circ \pi_1(\Gamma, \phi(\Gamma))$. Lemma 3.9.3 of Van Der Vaart and Wellner (1996) implies

$$\pi'_{\Gamma_0}(\gamma) = \phi'_{\pi_2 \circ \pi_1(\Gamma_0, \phi(\Gamma_0))} \circ \pi'_{2, \pi_1(\Gamma_0, \phi(\Gamma_0))} \circ \pi'_{1, (\Gamma_0, \phi(\Gamma_0))}(\gamma, \phi'_{\Gamma_0}(\gamma)) \quad (\text{C.4})$$

Next, the map π_1 is Hadamard differentiable at $\Lambda_0 = (\Lambda_{10}, \Lambda_{20}) \in l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T})$ tangentially to $l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T})$ with derivative at Λ_0 in $\lambda = (\lambda_1, \lambda_2)$ in $l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T})$ given by

$$\pi'_{1, \Lambda_0}(\lambda) = \lambda_1 - \lambda_2 \quad (\text{C.5})$$

Next, the map π_2 is Hadamard differentiable at Θ_0 tangentially to $l^\infty(\mathcal{UT})$ with derivative at Θ_0 in $\theta \in l^\infty(\mathcal{UT})$ given by

$$\pi'_{2, \Theta_0}(\theta) = 2\Theta_0\theta \quad (\text{C.6})$$

Consider any sequence $t_k > 0$ and $\Theta_k \in l^\infty(\mathcal{UT})$ and for $k = 1, 2, 3, \dots$ $t_k \downarrow 0$ and

$$\theta_k = \frac{\Theta_k - \Theta_0}{t_k} \rightarrow \theta \in l^\infty(\mathcal{UT}) \text{ as } k \rightarrow \infty$$

Then,

$$\begin{aligned}\frac{\pi_2(\Theta_k) - \pi_2(\Theta_0)}{t_k} - \pi'_{2,\Theta_0}(\theta) &= \frac{(\Theta_0 + t_k\theta_k)^2 - \Theta_0^2}{t_k} - 2\Theta_0\theta \\ &= 2\Theta_0(\theta_k - \theta) + \theta_k^2 t_k \rightarrow 0 \text{ as } k \rightarrow \infty\end{aligned}$$

which shows the result. And the main result follows from Lemma 7 and Equations C.4, C.5, and C.6. □

Proof of Theorem B.1

For the first part of the result,

$$\begin{aligned}\sqrt{n} \left(\hat{Var}(Y|T=t) - Var(Y|T=t) \right) &= \sqrt{n} \left(\frac{1}{S} \sum_{s=1}^S \left(\hat{Q}_{Y|T}(\tau_s|t) - \hat{E}[Y|T=t] \right)^2 - \pi(\hat{Q}_{Y|T}) \right) \\ &\quad + \sqrt{n}(\pi(\hat{Q}_{Y|T}) - \pi(Q_{Y|T})) \\ &= \pi'_{Q_{Y|T}} \sqrt{n}(\hat{Q}_{Y|T} - Q_{Y|T}) + o_p(1)\end{aligned}$$

which holds uniformly in t as long as S is large enough, e.g. $S = Cn^{1/2+\delta}$ for $C > 0$ and $\delta > 0$, and by Lemma 6. This implies the result. The second and third parts of the result hold using the same arguments.

C.8 Proof of Theorem B.2

The result follows immediately from Theorem 2

D Tables and Figures

Table 1: Summary Statistics

	Q1	Q2	Q3	Q4	All
Parents' Income (1000s)	32.44 (0.288)	51.46 (0.142)	68.04 (0.181)	107.55 (1.265)	64.87 (0.56)
Child's Income (1000s)	46.37 (0.944)	64.24 (1.486)	74.97 (1.465)	96.97 (2.349)	70.64 (0.873)
Head White	0.72 (0.015)	0.93 (0.008)	0.95 (0.007)	0.96 (0.006)	0.89 (0.005)
Head Non-White	0.28 (0.015)	0.07 (0.008)	0.05 (0.007)	0.04 (0.006)	0.11 (0.005)
Child Male	0.48 (0.016)	0.47 (0.016)	0.48 (0.016)	0.5 (0.016)	0.48 (0.008)
Head Male	0.77 (0.014)	0.95 (0.007)	0.97 (0.006)	0.99 (0.003)	0.92 (0.004)
Year Born	1970.18 (0.33)	1969.84 (0.319)	1970.19 (0.325)	1968.81 (0.353)	1969.75 (0.166)
Head Veteran	0.25 (0.014)	0.38 (0.016)	0.47 (0.016)	0.48 (0.016)	0.4 (0.008)
Head Less than HS	0.36 (0.016)	0.2 (0.013)	0.09 (0.009)	0.05 (0.007)	0.17 (0.006)
Head HS	0.57 (0.016)	0.63 (0.016)	0.58 (0.016)	0.38 (0.016)	0.54 (0.008)
Head College	0.06 (0.008)	0.17 (0.012)	0.33 (0.015)	0.57 (0.016)	0.28 (0.007)
Cutoff (1000s)	44.17	59.16	78.16	434.44	
N	932	932	931	932	3727

Notes: Summary statistics for the main dataset used in the paper. Each column provides average values of available variables by parents' income quartile. Standard errors are given in parentheses beneath the average. The row "Cutoff" is the maximum value of parents' income in that quartile (i.e. the dividing line between parents' income across two columns).

Sources: Panel Study of Income Dynamics, as described in text

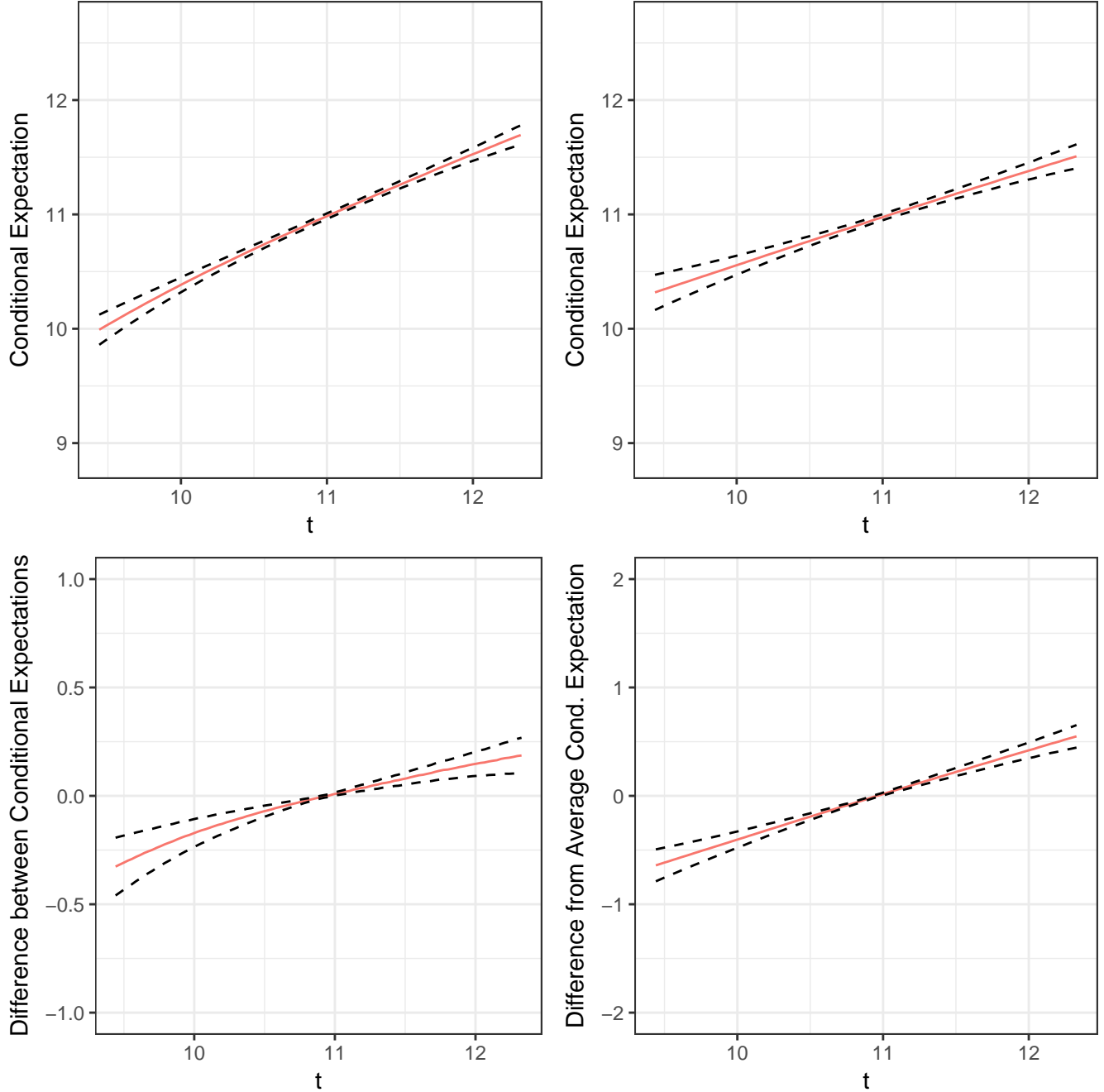
Table 2: Intergenerational Elasticity (IGE) Estimates

	<i>Dependent variable:</i>			
	Log Child's Income			
	(1)	(2)	(3)	(4)
Log Parents' Income	0.603*** (0.023)	0.562*** (0.024)	0.548*** (0.024)	0.443*** (0.027)
Head Non-White		-0.262*** (0.040)	-0.269*** (0.040)	-0.253*** (0.039)
Male		0.027 (0.019)	0.025 (0.019)	0.026 (0.019)
Head Male		-0.065 (0.043)	-0.043 (0.043)	-0.011 (0.043)
Year Born			-0.009*** (0.001)	-0.012*** (0.001)
Head Veteran				0.0002 (0.020)
Head Less Than HS Educ.				-0.215*** (0.029)
Head College Educ.				0.101*** (0.024)
Constant	4.348*** (0.251)	4.876*** (0.258)	22.246*** (1.901)	30.329*** (2.118)

Notes: Results come from regressions of the log of child's income on the log of parents' income and additional controls using the full sample of 3,727 observations. Standard errors are heteroskedasticity robust.

Sources: Panel Study of Income Dynamics, as described in text

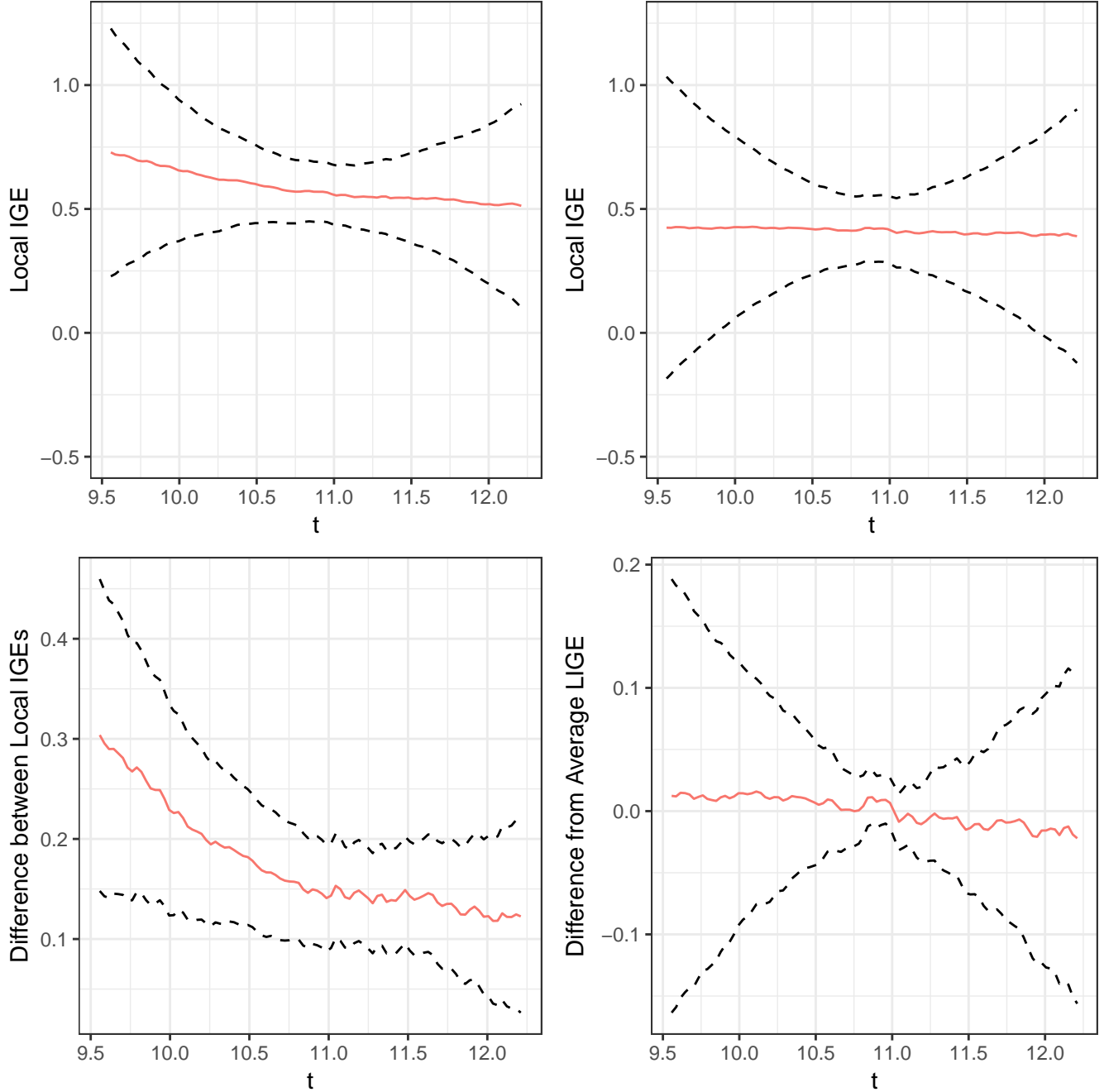
Figure 1: Expected Child's Income Conditional on Parents' Income



Notes: The top left panel plots average child's income as a function of parents' income with no adjustments for other covariates. The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.

Sources: Panel Study of Income Dynamics, as described in text

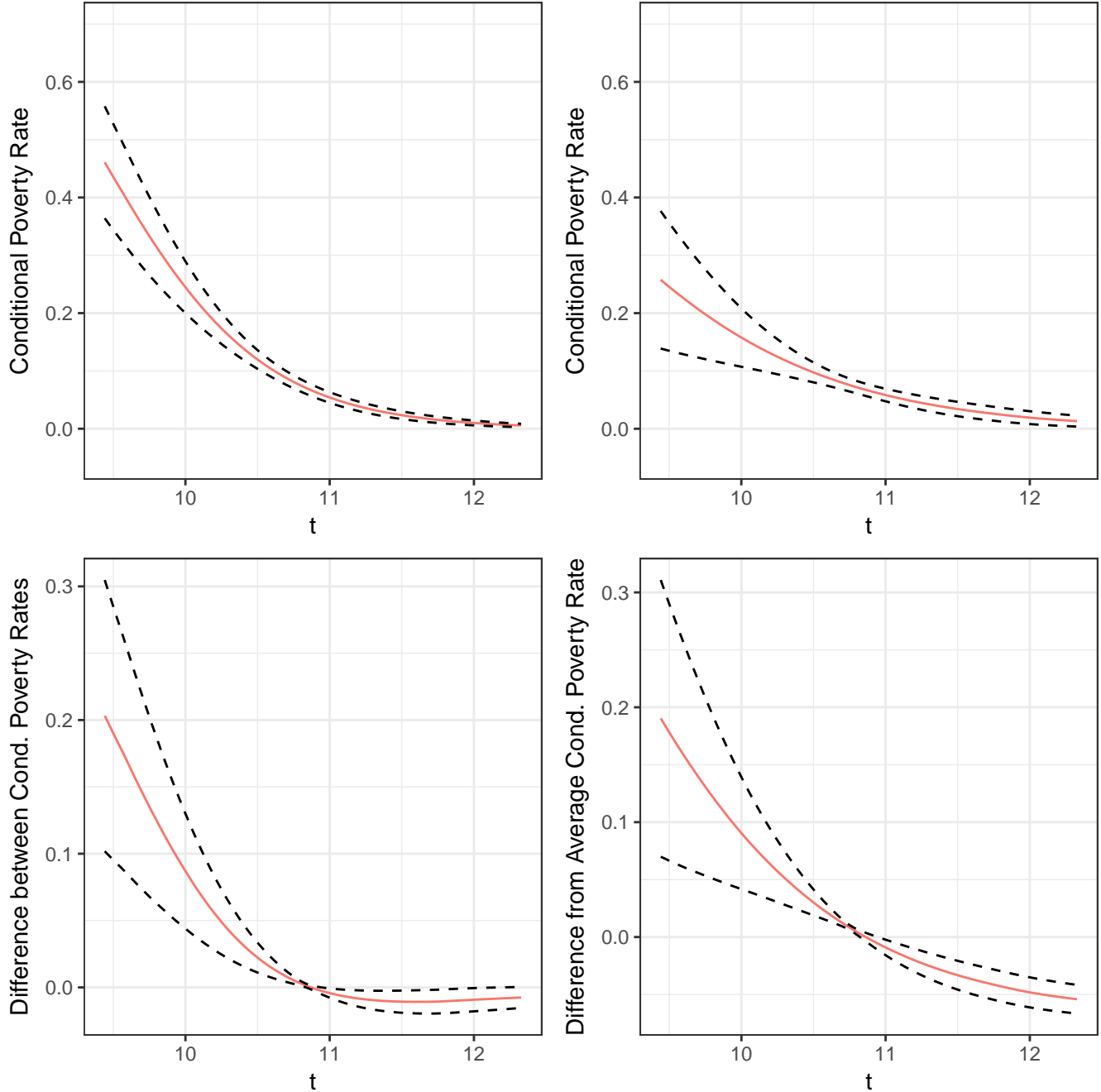
Figure 2: Local Intergenerational Elasticities (IGEs)



Notes: The top left panel plots the Local IGE as a function of parents' income with no adjustments for other covariates. Here we set $\delta = 0.1$, where δ is a fixed bandwidth as described in the main text. The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.

Sources: Panel Study of Income Dynamics, as described in text

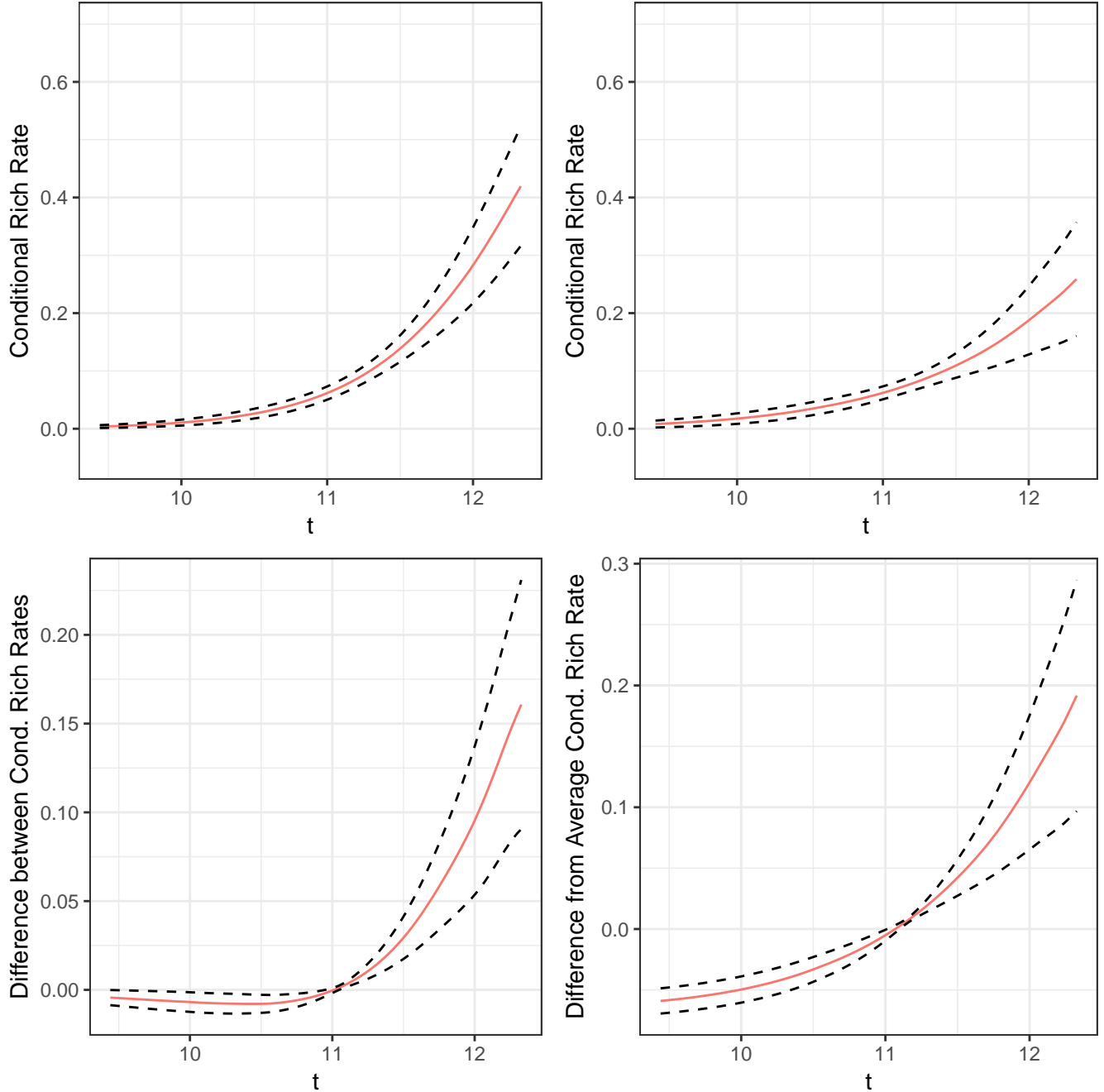
Figure 3: Fraction of Children below the Poverty Line



Notes: The top left panel plots the fraction of children below the poverty line as a function of parents' income with no adjustments for other covariates. The poverty line is set to be \$22,113 which is the poverty line for a family with two adults and two children in 2010. The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.

Sources: Panel Study of Income Dynamics, as described in text

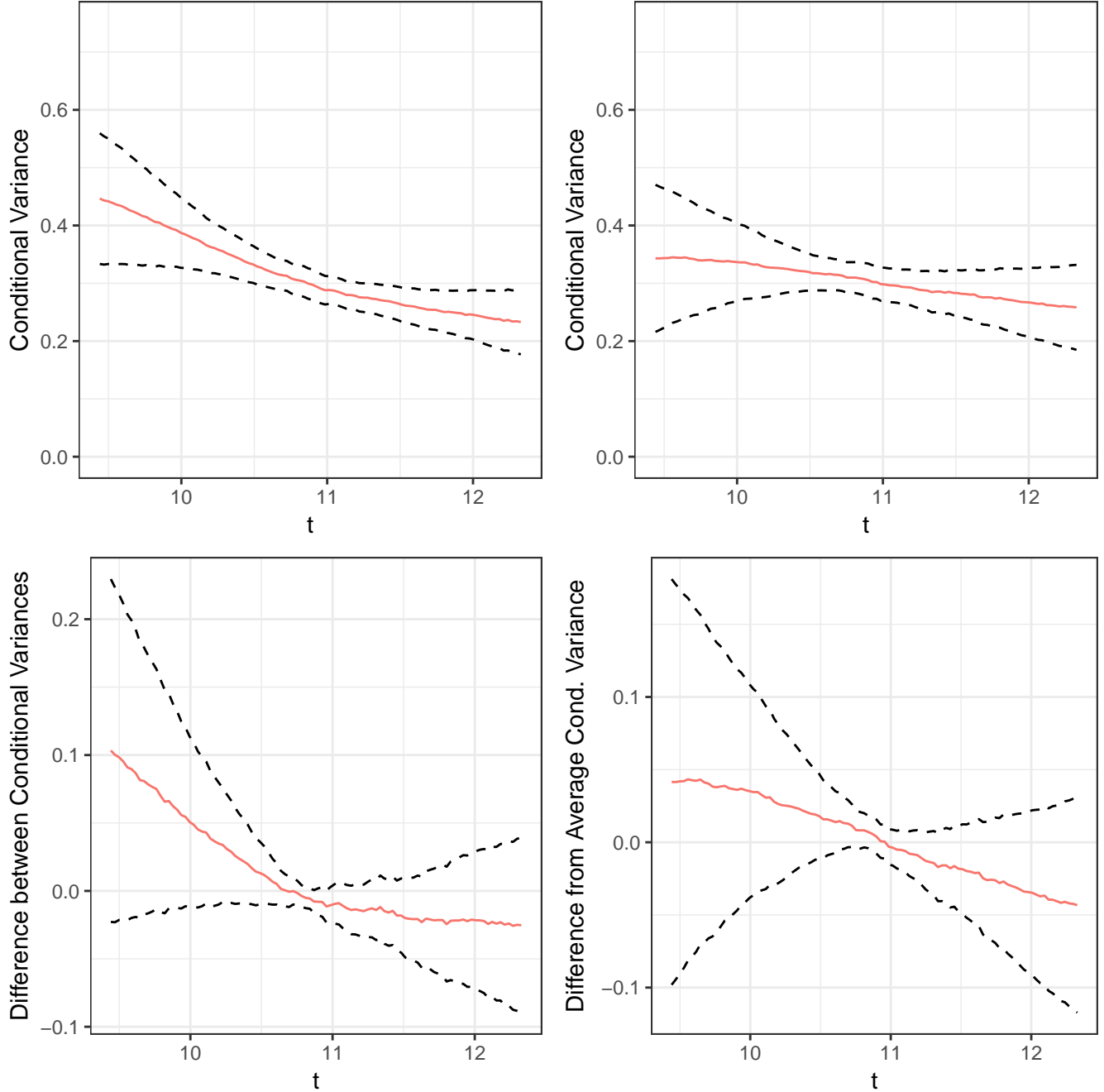
Figure 4: Fraction of “Rich” Children



Notes: The top left panel plots the fraction of “rich” children as a function of parents’ income with no adjustments for other covariates where “rich” is defined as having income above the 90th percentile of income in the U.S. in 2010 which is \$132,923. The top right panel adjusts for differences in the covariates family head’s race, family head’s gender, gender of child, child’s birth year, family head’s veteran status, and family head’s education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents’ income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.

Sources: Panel Study of Income Dynamics, as described in text

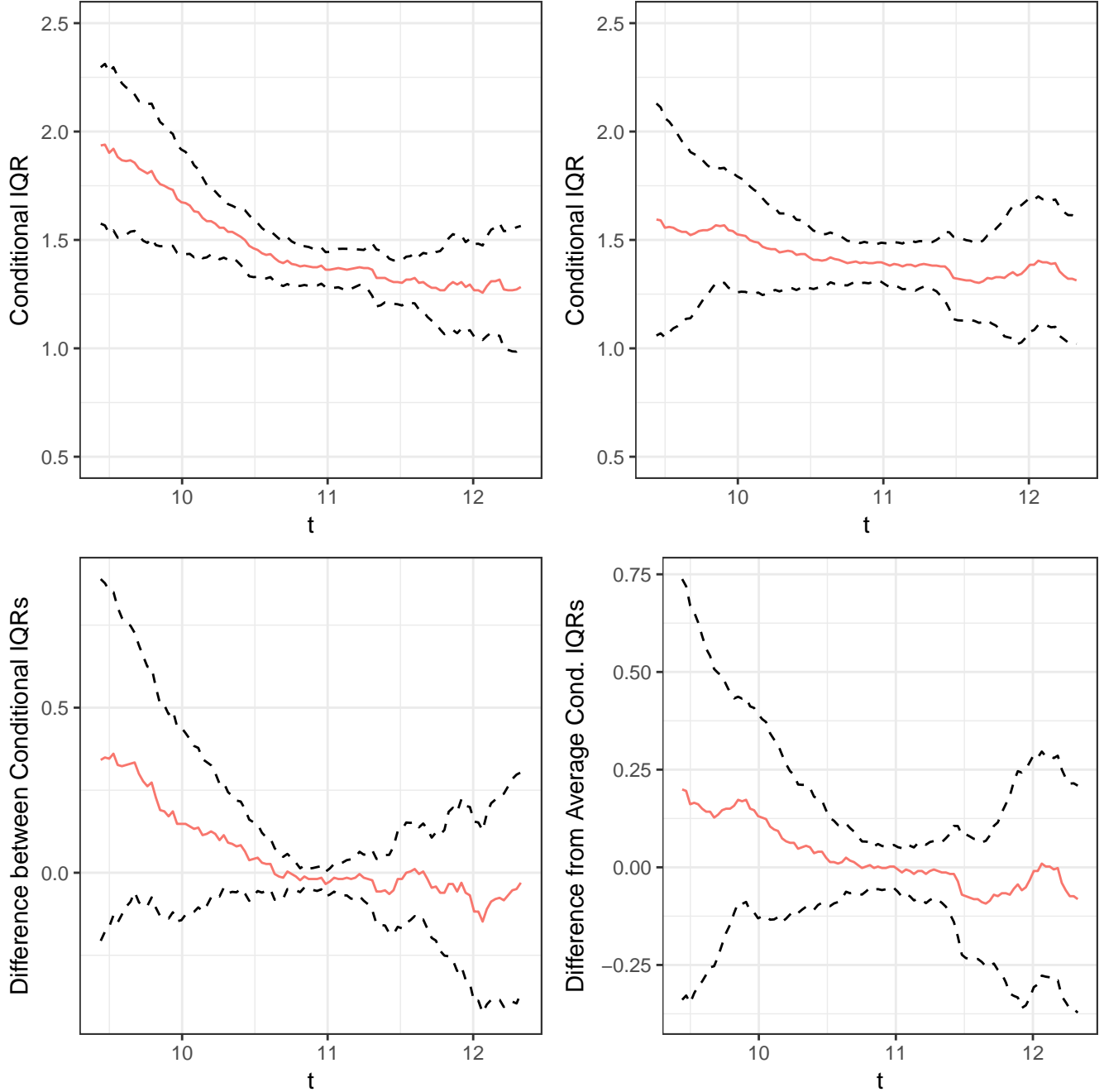
Figure 5: Variance of Child's Income Conditional on Parents' Income



Notes: The top left panel plots the variance of child's income as a function of parents' income with no adjustments for other covariates. The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.

Sources: Panel Study of Income Dynamics, as described in text

Figure 6: Interquantile Ranges



Notes: The top left panel plots the Interquantile Range as a function of parents' income with no adjustments for other covariates for $\tau_1 = 0.9$ and $\tau_2 = 0.1$ (these are the values of τ_1 and τ_2 used in each panel). The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, and family head's education (dummy variables for less than high school degree, high school degree but less than college degree, and college degree or more). The bottom left panel plots the difference between the estimates that do not adjust for covariates and that do adjust for covariates (i.e. the difference between the top left and top right panels as a function of parents' income). The bottom right panel plots the difference between the results that adjust for covariates and the average over t of the same results, as discussed in the text. In each panel, the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations as described in the text.

Sources: Panel Study of Income Dynamics, as described in text