Distributional Effects of a Continuous Treatment with an Application on Intergenerational Mobility*

Brantly Callaway† Weige Huang‡

October 8, 2018

Abstract

This paper considers the effect of a continuous treatment on the entire distribution of outcomes after adjusting for differences in the distribution of covariates across different levels of the treatment. Our methodology encompasses dose response functions, counterfactual distributions, and "distributional policy effects" depending on the assumptions invoked by the researcher. We propose a three-step estimator that consists of (i) estimating the distribution of the outcome conditional on the treatment and other covariates using quantile regression; (ii) for each value of the treatment, averaging over a counterfactual distribution of the covariates holding the treatment fixed; (iii) manipulating the counterfactual distribution into a parameter of interest. 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 the treatment. We use our method to study intergenerational income mobility where we consider distributional effects of parents' income on child's income such as (i) the fraction of children with income below the poverty line, (ii) the variance of child's income, and (iii) the inter-quantile range of child's income – all as a function of parents' income.

Keywords: Counterfactual Distribution, Dose Response Function, Distributional Policy Effect, Continuous Treatment, Treatment Effects, Quantile Regression, Intergenerational Income Mobility

JEL Codes: C21, J62

*We thank Afrouz Azadikhah-Jahromi, David Kaplan, Martin Lopez-Daneri, Catherine Maclean, Irina Murtazashvili, 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. This paper replaces "Intergenerational Income Mobility: Counterfactual Distributions with a Continuous Treatment." 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

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

1 Introduction

Researchers in economics often consider the effect of one variable (a treatment) on an outcome while perhaps adjusting for differences in other variables that are related to outcomes and that are distributed differently across different values of the treatment. While the case of a binary treatment has received much attention in the literature (see, for example, the review of Imbens and Wooldridge (2009)), this paper develops new methods for the case with a continuous treatment which has received considerably less attention.

Methods for dealing with a continuous treatment are likely to be of interest to empirical researchers across a variety of areas. To give some examples of applications with a continuous treatment, Imbens, Rubin, and Sacerdote (2001) study the effect of unearned income on labor market outcomes; Galvao and Wang (2015) consider an application on the effect of mother's weight gain during pregnancy on child's birth weight; Jasova, Mendicino, and Supera (2018) study the effect of the term structure of debt on banks' lending behavior; and local labor markets approaches common in labor economics often involve a continuous treatment (for example, Autor, Dorn, and Hanson 2013; Acemoglu and Restrepo 2017; Collins and Niemesh 2017).

This paper proposes simple, but flexible, semiparametric estimators of distributional effects[1] of a continuous treatment while adjusting for differences in the distribution of covariates across different levels of the treatment. Our procedure requires three steps. First, we estimate the distribution of the outcome conditional on the treatment and other observed characteristics by inverting quantile regression (Koenker and Bassett Jr 1978; Koenker 2005; Chernozhukov, Fernandez-Val, and Melly 2013) estimates of the conditional quantiles. The second step is to estimate the "counterfactual distribution" which involves integrating over the first step estimates while changing the distribution of observable characteristics. This step obtains the entire distribution of the outcome

[1]We use the term distributional effects broadly here. It encompasses counterfactual distributions, dose response functions, treatment effects, and distributional policy effects with a continuous treatment depending on the assumptions invoked by the researcher. Dose response functions and treatment effects are defined in terms of potential outcomes, and we consider these under the assumption of unconfoundedness. Counterfactual distributions are similar, but do not rely on potential outcomes notation or the assumption of unconfoundedness. The counterfactual distributions that we consider fix the return to characteristics (for a particular value of the treatment) but change the distribution of the characteristics. Distributional policy effects compare (functionals of) the counterfactual distribution to (functionals of) the observed distribution of outcomes conditional on the treatment. They are closely related to composition effects in the literature on decompositions.

as a function of the continuous treatment after adjusting for differences in the distribution of covariates across different levels of the treatment. Finally, parameters of interest such as measures of the spread of the outcome and the probability of having a very low outcome (e.g. child's income being below the poverty line) are obtained as functions of the counterfactual distribution.

The literature on continuous treatment effects includes Hirano and Imbens (2004), Flores (2007), Florens, Heckman, Meghir, and Vytlacil (2008), Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), Galvao and Wang (2015), and Kennedy, Ma, McHugh, and Small (2016). Within this literature, the only paper that we are aware of that looks at distributional parameters is Galvao and Wang (2015) which proposes a weighting estimator of quantile dose response functions with a continuous treatment under the assumption of unconfoundedness. Our approach is different in that our estimators are based on first step quantile regression and do not require estimating conditional densities in the first step.[2] There are trade-offs to using quantile regression relative to weighting estimators based on conditional densities. Quantile regression imposes stronger parametric assumptions than nonparametrically estimating conditional densities, though it is much simpler to implement in practice; on the other hand, quantile regression is much more flexible (though perhaps somewhat harder to implement) than assuming a fully parametric model for a conditional density.[3]

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)) and results on first step quantile regression estimators (Chernozhukov, Fernandez-Val, and Melly 2013). We show that the limiting processes can be approximated using the empirical bootstrap. In the context of intergenerational mobility, these results allow us to test functional hypotheses about income mobility such as (1) whether adjusting

[2]Another 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.

[3]Our paper is also related to the literature on decompositions with a continuous treatment. Ñopo (2008) and Ulrick (2012) provide decompositions for the mean with a continuous treatment. Ao, Calonico, and Lee (2017) consider decompositions with a multi-valued discrete treatment. Bowles and Gintis (2002), Groves (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.

for covariates has any effect on any particular parameter (e.g. the percentage of children with income below the poverty line as a function of parents' income), or (2) whether parameters of interest are the same at all values of parents' income (e.g. the variance of child's income).

We use our method to study the effect of parents' income on child's income. 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 child'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 \$33,800 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 \$33,800, or (ii) the variance of these individuals' income is high implying that some of them have much higher incomes than \$33,800 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.

From a methodological perspective, a key 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), 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 study counterfactual distributions with a continuous treatment.

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 fraction of children whose incomes are below the poverty line, (3) the variance of child's income, (4) the inter-quantile range of child's income, 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.

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 higher income on average 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 last three results are similar to existing results using transition matrices though our approach does not require specifying cutoffs of the continuous treatment and provides a straightforward way to incorporate adjusting for covariates.

A second motivation of our paper is to look at the role that background characteristics play in the transmission of income across generations. We 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 quarter 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.

2 Parameters of Interest

This section develops several distributional parameters of interest for the case of a continuous treatment. We are motivated by our application on intergenerational income mobility, but these parameters are likely to be of interest in other applications as well. This section also distinguishes between several classes of parameters: treatment effects, dose response functions, counterfactual distributions, and distributional policy effects. The differences are based primarily on (i) the particular application and (ii) whether or not the researcher wishes to invoke the assumption of unconfoundedness.

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). In the application, Y is the log of child's income, T is the log of parents' income and X are additional covariates such as parents' education, child's birth year, gender, and race.

2.1 Identification

We use the following notation. Let Y denote an individual's outcome, T denote an individual's level of treatment, and X denote a $k \times 1$ vector of covariates. We let $\mathcal{Y}$, $\mathcal{T}$, and $\mathcal{X}$ denote the supports of Y, T, and X. Next, let $Y(t)$ denote an individual's potential outcome – the outcome that would occur for an individual if they experienced treatment level t. We consider the following assumptions.

Assumption 1 (Unconfoundedness)**.**

$$Y(t) \perp\!\!\!\perp T | X$$

Assumption 2 (Common Support)**.**

$f_{T|X}$ is uniformly bounded away from 0 and ∞ on $\mathcal{T}\mathcal{X}$.

Assumption 1 says that, conditional on covariates X, treatment is as good as randomly assigned. Unconfoundedness is also known as selection on observables or ignorability. Some version of Assumption 1 is invoked in much of the literature on continuous treatment effects (e.g. Hirano and Imbens (2004), Flores (2007), Flores, Flores-Lagunes, Gonzalez, and Neumann (2012), and Galvao and Wang (2015)). It is very closely related to the unconfoundedness assumption in the literature with a binary treatment (e.g. Rosenbaum and Rubin (1983), Heckman, Ichimura, and Todd (1997), Hirano, Imbens, and Ridder (2003), and Imbens and Wooldridge (2009)). Assumption 2 imposes a common support assumption. It says that, for all values of the covariates, there are individuals that experience each level of the treatment. This is a strong assumption. In the context of intergenerational mobility, for example, it requires that there be some very poor parents with very high education as well as some very rich parents with very low education. This type of assumption is common in the treatment effects literature though; and, in the case where a researcher is interested in effects at particular values of t, it could be weakened to hold only at those values of the treatment. Under Assumptions 1 and 2, it is straightforward to show that for any $t \in \mathcal{T}$,

$$P(Y(t) \leq y) = \int_{\mathcal{X}} F_{Y|T,X}(y|t,x)\, dF_X(x) \tag{2.1}$$

which says that the distribution of potential outcomes if all individuals were assigned the treatment level t can be obtained by integrating the distribution of Y conditional on X and T over the distribution of X for the entire population. In the continuous treatment effect literature, $P(Y(t) \leq y)$ is called the distribution dose response function. One can invert the distribution dose response function to obtain the quantile dose response function or one can integrate over the distribution to obtain the average dose response function. We discuss more parameters of interest in the next subsection.

Interestingly, even without Assumption 1, the term on the right hand side of Equation (2.1) has a useful interpretation. First, notice that the observed distribution of the outcome conditional on the treatment is given by:

$$F_{Y|T}(y|t) = \int_{\mathcal{X}} F_{Y|T,X}(y|t,x)\; dF_{X|T}(x|t) \tag{2.2}$$

that is, $F_{Y|T}(y|t)$ is the same as integrating the distribution of the outcome Y conditional on observed characteristics X and the treatment $T = t$ over the distribution of X conditional on $T = t$. One can also consider the *counterfactual distribution* of outcomes that individuals that experience treatment level t would experience if the returns to observed characteristics X were held constant but the distribution of covariates for individuals with treatment level t were manipulated to be the same as the distribution of covariates for the entire population. It is given by

$$F^C_{Y|T}(y|t) = \int_{\mathcal{X}} F_{Y|T,X}(y|t,x)\; \mathrm{d}F_X(x) \tag{2.3}$$

which is the same as Equation (2.1) and where we use the superscript C to indicate that it is a counterfactual distribution.[4] Notice that Equation (2.3) does not require Assumption 1 nor

[4]The counterfactual distribution mentioned above is not the only possible counterfactual distribution, though

does it require potential outcomes notation. We use the terminology counterfactual distribution throughout the rest of the paper; however, depending on the application, a researcher may wish to invoke Assumption 1 with the payoff being that the resulting parameters may be interpreted as causal effects.

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 2.3 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 2.3 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 functions[5] which are more challenging to estimate than the conditional distribution function above. Relative to weighting estimators, our approach can be seen as a "regression-adjustment" approach (see Wooldridge (2010)[Section 21.3.2]) to estimating distributional effects with a continuous treatment.

In the context of intergenerational mobility, the counterfactual distribution of child's income 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

it is the most common. Other manipulations of the distribution of the covariates are possible (see the discussion in Rothe (2010) and Chernozhukov, Fernandez-Val, and Melly (2013)).

[5]With 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.

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.

2.2 Parameters of Interest

The observed distribution $F_{Y|T}$ of the outcome conditional on the treatment and the counterfactual distribution $F^C_{Y|T}$ contain much useful information, but they suffer from being difficult to interpret or plot directly. In particular, they are both indexed by y and t which makes plots that vary both y and t three dimensional and difficult to easily interpret. Instead, we focus on estimating functionals of $F_{Y|T}$ and $F^C_{Y|T}$. This section covers these functionals.

Fraction of Individuals with "Low" Outcomes as a Function of the Treatment

The first parameter that we consider is the fraction of individuals whose outcome falls below a particular cutoff y_p as a function of the treatment variable. This is a particularly interesting parameter in the context of intergenerational income mobility. Set y_p equal to the poverty line. Then, this parameter is the fraction of children whose permanent income falls below the poverty line as a function of parents' income. This is given by

$$F_{Y|T}(y_p|t) \qquad \text{and} \qquad F^C_{Y|T}(y_p|t)$$

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.

For intergenerational income mobility, 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

$$1 - F_{Y|T}(y_R|t) \qquad \text{and} \qquad 1 - F^C_{Y|T}(y_R|t)$$

coming from the observed distribution and the counterfactual distribution, respectively.

Quantiles of the Outcome as a Function of the Treatment

One can obtain the quantiles of the outcome as a function of the treatment from the observed distribution and counterfactual distributions. For some $\tau \in (0,1)$, these are given by

$$Q_{Y|T}(\tau|t) = \inf\{y : F_{Y|T}(y|t) \geq \tau\} \qquad \text{and} \qquad Q^C_{Y|T}(\tau|t) = \inf\{y : F^C_{Y|T}(y|t) \geq \tau\}$$

for the quantiles of the observed distribution and the quantiles of the counterfactual distribution, both as a function of the treatment. Under Assumption 1, $Q^C_{Y|T}$ is called the quantile dose response function in Galvao and Wang (2015). The quantiles are also useful inputs into the remaining parameters of interest.

Average Outcome as a Function of the Treatment

The next parameter that we consider is the average outcome as a function of the treatment which is given by

$$E[Y|T=t] = \int_0^1 Q_{Y|T}(\tau|t)\ \mathrm{d}\tau \qquad \text{and} \qquad E^C[Y|T=t] = \int_0^1 Q^C_{Y|T}(\tau|t)\ \mathrm{d}\tau$$

where these depend on the observed distribution and counterfactual distribution, respectively.[6] 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.

Measures of Spread of the Outcome as a Function of the Treatment

Because our method obtains the entire observed distribution and counterfactual distribution

[6] In practice, we trim out the uppermost and lowermost quantiles so that the integration is from ϵ to $1-\epsilon$ (for some small positive ϵ) though it is likely to be possible to integrate from 0 to 1 under some additional conditions as in Bhattacharya (2007), Barrett and Donald (2009), and Donald, Hsu, and Barrett (2012).

of the outcome conditional on the treatment, we can study other features of these distributions than just their mean. In this section, we consider the variance of the outcome conditional on the treatment and the inter-quantile range of the outcome conditional on the treatment. For intergenerational mobility, these give measures of the spread of child's income conditional on parents' income.

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 the outcome as a function of the treatment.[7] It is given by

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

and

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

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

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

[7] In practice, for the variance, we trim out the uppermost and lowermost quantiles just like we did for the mean. See the discussion in Footnote 6.

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.[8]

Distributional Policy Effects

Another interesting class of parameters to consider are those that determine how (functionals) of the distribution of outcomes change as a result of adjusting for covariates. Rothe (2010) terms these types of parameters "distributional policy effects," and they are given by the difference between (functionals) of $F_{Y|T}$ and $F^C_{Y|T}$.[9] We denote these by

$$\Delta_k(t) = \Gamma_k(F_{Y|T}) - \Gamma_k(F^C_{Y|T})$$

where k indexes some particular parameter and Γ_k denotes the functional that transforms a conditional distribution into the parameter of interest. For example, to examine the role that adjusting for differences in covariates plays in terms of the fraction of children with income below the poverty line, we can consider the parameter

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

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. Similarly, 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

[8] There are other parameters as well that we could consider given that the observed and counterfactual distributions are identified. For example, Barrett and Donald (2009) consider several varieties of Lorenz curves and Gini coefficients that would be of interest in the case where a researcher is interested in inequality as a function of the treatment.

[9] Distributional policy effects are also closely related to the composition effect in aggregate decompositions.

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.

Treatment Effects

If one imposes Assumption 1, then our results are closely related to treatment effects. 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. Setting $t = t_{0.75}$ and $t' = t_{0.25}$ which represent the 75th percentile and 25th percentile of the treatment, 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. Similarly,

$$DTE(t,t';y_p) = F^C_{Y|T}(y_p|t) - F^C_{Y|T}(y_p|t')$$

is how much the fraction of individual's with income below the poverty line changes for t relative to t'.

2.3 Testing if Parameters Depend on the Level of the Treatment

Each of the parameters mentioned above can be considered as a function of the treatment t. As a final step in our analysis, we are interested in testing whether the treatment 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

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

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

$$H_0 : R_\theta(t) = 0 \qquad \text{for all } t \in \mathcal{T} \tag{2.4}$$

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' income. 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 the outcome Y conditional on the treatment T and possibly other observed characteristics X using quantile regression to obtain the conditional quantiles and then inverting to obtain the conditional distribution. 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 $F^C_{X|T} = F_X$; 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.

3.1 Step 1: Estimating the Conditional Distribution

We estimate the conditional distribution function $F_{Y|T,X}$ using quantile regression (Koenker and Bassett Jr 1978; Koenker 2005; Chernozhukov, Fernandez-Val, and Melly 2013).[10] We make the following assumptions

[10] In the Supplementary Appendix, we consider an alternative approach based on first step distribution regression (Foresi and Peracchi 1995; Chernozhukov, Fernandez-Val, and Melly 2013).

Assumption 3. *For all $\tau \in \mathcal{T}$*

$$Q_{Y|T,X}(\tau|t,x) = P_1(t,x)'\alpha(\tau)$$

where $P_1(t,x)$ are functions of t and x (e.g. the leading special case is that $P_1(t,x)$ is given by the $(k+1) \times 1$ vector $(t,x')'$ though it can also include interactions, higher order terms, etc.) and α is a $dim(P_1(t,x)) \times 1$ vector of parameters indexed by τ.

Assumption 4. *For all $\tau \in \mathcal{T}$*

$$Q_{Y|T}(\tau|t) = P_2(t)'\beta(\tau)$$

where $P_2(t)$ are functions of t and $\beta(\tau)$ is a $dim(P_2(t)) \times 1$ vector of parameters indexed by τ.

Assumptions 3 and 4 impose that the conditional quantiles are linear in parameters and can be estimated using standard quantile regression techniques. With the conditional quantiles in hand, the conditional distribution can be obtained by inverting the conditional quantiles. To implement the quantile regression estimator, we estimate the conditional quantiles over a fine, equally-spaced grid of S possible values for τ satisfying $0 < \tau_1 < \cdots < \tau_S < 1$. The estimated conditional quantiles are given by

$$\hat{Q}_{Y|T,X}(\tau|t,x) = P_1(t,x)'\hat{\alpha}(\tau) \qquad \text{and} \qquad \hat{Q}_{Y|T}(\tau|t) = P_2(t)'\hat{\beta}(\tau)$$

and which can each be inverted to obtain the conditional distributions by

$$\hat{F}_{Y|T,X}(y|t,x) = \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\{\hat{Q}_{Y|T,X}(\tau_s|t,x) \leq y\} \qquad \text{and} \qquad \hat{F}_{Y|T}(y|t) = \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\{\hat{Q}_{Y|T}(\tau_s|t) \leq y\}$$

$\hat{F}_{Y|T}$, given above, is the observed distribution which is one of our objects of interest; however, we still need to manipulate $\hat{F}_{Y|T,X}$ to be the counterfactual distribution of interest.

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 $F^C_{Y|T}(y|t)$ amounts to averaging over X while holding t fixed. That is,

$$\hat{F}^C_{Y|T}(y|t) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}_{Y|T,X}(y|t, X_i)$$

which is the same as replacing the population distribution function in Equation 2.3 with the sample distribution function. We plug in the estimates $\hat{F}_{Y|T}$ and $\hat{F}^C_{Y|T}$ below to obtain estimates of particular parameters of interest.

3.3 Step 3: Estimating Parameters of Interest

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

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

Estimating quantiles of the outcome conditional on the treatment can also be obtained simply by plugging in to the results in Section 2:

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

which simply inverts the counterfactual distribution of outcomes. Next, we can estimate $E[Y|T = t]$ by.

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

where $0 < \tau_1 < \tau_2 < \cdots < \tau_S < 1$ is the grid of values of τ given above. 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

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

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

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

3.4 Asymptotic Theory

This section develops asymptotic theory and inference procedures for the parameters discussed in Section 2. Our inference results are uniformly valid in the treatment T, and we derive the joint limiting distributions of parameters that depend on the observed distribution $F_{Y|T}$ and on the counterfactual distribution $F^{C}_{Y|T}$. 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 the treatment, (2) whether any parameter of interest (such as the variance or inter-quantile range) of the outcome is constant across different values of the treatment, 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) and, in particular, they build off the theoretical results on first step quantile regression in Chernozhukov, Fernandez-Val, and Melly (2013).[11] 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

[11] This is similar to other papers that also use quantile regression as a first step estimator and build off the same results in other contexts; e.g. Melly and Santangelo (2015) and Wuthrich (2015).

empirical bootstrap is valid for conducting inference – both uniformly and pointwise. All proofs are contained in the Appendix. We make the following assumption,

Assumption 5. *(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 quantile 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

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

denote the empirical process of the counterfactual distribution of the outcome conditional on the treatment. 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 the outcome conditional on the treatment.

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 Assumptions 2 to 5 and A.1 (given in Appendix A)*

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

in the space $\mathbb{S}$ *where* $(\mathbb{V}_{Y|T}, \mathbb{V}^C_{Y|T})$ *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* $\mathbb{V}^C_{Y|T}(y|t) = \int_{\mathcal{X}} \mathbb{G}_{Y|T,X}(y|t,x)\, \mathrm{d}F_X(x) + \int_{\mathcal{X}} F_{Y|T,X}(y|t,x)\, \mathrm{d}\mathbb{G}_X(x)$ *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. It essentially follows from the results in Chernozhukov, Fernandez-Val, and Melly (2013) with relatively small differences related to establishing the joint limiting process. It should also be noted that our results hold uniformly in the treatment T though this does not require major changes in the theory. 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 considers the joint limiting process of the observed distribution and the counterfactual distribution which allows one to consider uniform inference on the *difference* between particular parameters under the observed distribution and counterfactual distribution. It will also be important for testing whether or not a particular parameter changes across different values of the treatment.

The next corollary provides a general result for the limiting process of Hadamard differentiable functions of $F_{Y|T}$ and $F^C_{Y|T}$. This result covers all of the parameters of interest in Section 2.

Corollary 1. *Let $\mathbb{D} = l^\infty(\mathcal{YT})$ and consider the Hadamard differentiable map $\Gamma : \mathbb{D}_\Gamma \subset \mathbb{D} \mapsto l^\infty(\mathcal{T})$ with derivative Γ'_γ for $\gamma \in \mathbb{D}$. Let $\hat{G}_T(t) = \sqrt{n}(\Gamma(\hat{F}_{Y|T}(\cdot|t)) - \Gamma(F_{Y|T}(\cdot|t)))$ and $\hat{G}^C_T(t) = \sqrt{n}(\Gamma(\hat{F}^C_{Y|T}(\cdot|t)) - \Gamma(F^C_{Y|T}(\cdot|t)))$. Then,*

$$(\hat{G}_T(t), \hat{G}^C_T(t)) \rightsquigarrow (\Gamma'_{F_{Y|T}}, \Gamma'_{F^C_{Y|T}})$$

in the space $l^\infty(\mathcal{T})^2$.

In the Supplementary Appendix we show that each of the parameters that we consider in the paper is indeed Hadamard differentiable and give explicit expressions for each term in Corollary 1 (which depend on the particular parameter of interest). In addition, given the results in Theorem 1 and Corollary 1, the validity of the empirical bootstrap for conducting uniform inference follows using well known arguments (for example, Van Der Vaart and Wellner (1996) and Kosorok (2007)). 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 $\theta(t)$. In particular, we can 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

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

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

$$\hat{\Sigma}(t) = \frac{q_{0.75}(t) - q_{0.25}(t)}{z_{0.75}(t) - z_{0.25}(t)} \tag{3.1}$$

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 $b = 1, \ldots, B$ 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 boostrapped sample. Then, setting $\hat{c}_{1-\alpha}$ to be the $(1-\alpha)$ quantile of $\{\hat{c}_b : 1 \le b \le B\}$ implies that $\hat{C}_\theta(t)$ asymptotically covers $\theta(t)$ for all values $t \in \mathcal{T}$ with probability $(1-\alpha)$.

We provide the theoretical justification for the above procedure in the Supplementary Appendix. Finally, and using similar arguments as above, one can establish the limiting process and prove the validity of the empirical bootstrap for testing whether or not parameters depend on the value of the treatment as in Section 2.3. We also provide the details for this procedure in the Supplementary Appendix.

4 Application on Intergenerational Income Mobility

4.1 Related Literature

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)). 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. Richey and Rosburg (2016) propose a similar counterfactual distribution to the one in the current paper in the context of a decomposition of intergenerational income mobility though they propose a first step distribution regression estimator and second step simulation estimator. The intergenerational elasticity (IGE), which is the slope coefficient from a regression of the log of child's income on the log of parents' income, has a long history in the intergenerational income mobility literature. But recent work has considered more complicated setups such as (1) transition matrices, (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, among other ideas (Jantti et al. 2006; Bhattacharya and Mazumder 2011; Murtazashvili 2012; Chetty, Hendren, Kline, and Saez 2014; Chetty et al. 2014; Murtazashvili, Liu, and Prokhorov 2015; An, Le, and Xiao 2017; Chetty et al. 2017; Collins and Wanamaker 2017; Kitagawa, Nybom, and Stuhler 2017).

Of these, transition matrices are most closely related to our approach and 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. A key advantage of our approach is that it does not require choosing cutoffs like this. Another advantage of our approach is that it is straightforward to include covariates in the analysis. A final distinction is that 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.

4.2 Data

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).[12] 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. Our sample consists of 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,

[12] 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.

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 3,630 child-parent 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 \$78,000. 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. 35% of family heads in the lowest quartile have less than a high school education. The corresponding quantities are 19%, 8%, 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 – 7% in the first quartile, 17% in the 2nd, 34% 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 0.609.[13] Adding demographic controls, as in specification (2) in the table, shrinks the estimated coefficient to 0.573. 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 as older individuals have more work experience; 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 0.452; so, here, adding additional controls reduced the estimated IGE by about 26%. The coefficients 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.

4.3 Main Results

Our main results are provided in Figures 1 to 5 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

[13] This 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).

not cover zero. We also impose that $Q_{Y|T,X}(\tau|t,x) = \alpha_1(\tau)t + x'\alpha_2(\tau)$ (and that X includes an intercept term) and that $Q_{Y|T}(\tau|t) = \beta_0(\tau) + \beta_1(\tau)t$, both for all values of τ.

Question 1: How much does adjusting for covariates matter?

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.[14] Average child's income as a function of parents' income is reported in Figure 1. 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 \$61,000 without adjusting for covariates and \$59,900 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.

As for our local IGE measures (the slope of average child's income as a function of parents' income), with or without adjusting for differences in covariates, they are roughly constant across all levels of parents' income. However, the level is quite different. Without adjusting for covariates, our local measure of IGE tends to be around 0.57 across all values of parents' income. Adjusting

[14] For the IGE, we take a numerical derivative of $E[Y|T]$ and $E^C[Y|T]$, respectively. For example, for the counterfactual IGE, we calculate $IGE^C(t) = \delta^{-1}(E^C[Y|T=t+\delta/2] - E^C[Y|T=t-\delta/2])$ for some small, fixed δ (we set this equal to 0.1 in practice). One difference between our results and most existing work is that our measure of the IGE is local, though Landersø and Heckman (2017) have considered a local IGE in previous work. However, the main departure here 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.

for covariates, the local IGE is around 0.42 across all value of parents' income. Taken together, these results suggest that, on average, the effect of increasing parents' income on child's income is roughly the same across all levels of parents' income and that taking into account differences in the distribution of covariates across different levels of parents' income tends to somewhat decrease the effect of parents' income – results that are in line with the existing literature on intergenerational income mobility. Next, we turn to looking at distributional effects of parents' income on child's income.

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 their 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 2. Without adjusting for covariates, 21.7% of children from families with incomes at the poverty line are estimated to have incomes below the poverty line themselves. After adjusting for covariates, only 16.3% are estimated to have incomes below the poverty line. At the median of parents' income, without adjusting for covariates 4.3% of children have income below the poverty line and slightly more, 4.7%, have income below the poverty line after adjusting for differences in observed characteristics (this difference is not statistically significant). For children of families in the 90th percentile of income, we estimate that only 1.0% have incomes below the poverty line without adjusting for covariates while 1.7% have incomes below the poverty line when we do adjust for covariates (this difference is not statistically significant). 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 2.0% of children from families at the poverty line, 7.4% of children from families at the median, and 26.7% 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 20.1%.

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 4 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 relatively 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 5 (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.

Treatment Effects

It seems unlikely that our estimates should be considered to be estimates of the causal effect of parents' income on child's income, but we briefly consider the relationship of our estimates to treatment effect estimates under Assumption 1.[15] We estimate that, on average, moving from the 25th percentile to the 75th percentile of parents' income increases child's income by 24.7 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 parents' income decreases the probability of a child's income being below the poverty line by 4.2 percentage points (a 59% reduction).

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 parents' 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.

[15]We suspect that, in the context of intergenerational income mobility, estimates of the average effect of moving from a low level of parents' income to a high level of parents' income using our approach are likely to overstate the causal 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 three 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.

5 Conclusion

This paper has developed new tools to study distributional effects of a continuous treatment. We proposed a straightforward three step procedure to estimate these distributional effects that is based on first step quantile regression. Our procedure is easy to implement in practical applications and more flexible than making distributional assumptions about the treatment or outcome.

We applied these methods 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 distributional effects with a continuous treatment or are interested in the causal effect of a continuous treatment under the assumption of unconfoundedness.

References

[1] Acemoglu, Daron and Pascual Restrepo. “Robots and jobs: Evidence from US labor markets”. Working Paper. 2017.

[2] An, Yonghong, Wang Le, and Ruli Xiao. “Your American Dream is not mine! A new approach to estimating intergenerational mobility elasticities”. Working Paper. 2017.

[3] Ao, Wallice, Sebastian Calonico, and Ying-Ying Lee. “Multivalued treatments and decomposition analysis: An application to the WIA program”. Working Paper. 2017.

[4] Autor, David, David Dorn, and Gordon H Hanson. “The China syndrome: Local labor market effects of import competition in the United States”. *American Economic Review* 103.6 (2013), pp. 2121–68.

[5] Barrett, Garry F and Stephen G Donald. “Statistical inference with generalized Gini indices of inequality, poverty, and welfare”. *Journal of Business & Economic Statistics* 27.1 (2009), pp. 1–17.

[6] Bhattacharya, Debopam. “Inference on inequality from household survey data”. *Journal of Econometrics* 137.2 (2007), pp. 674–707.

[7] 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.

[8] Black, Sandra E and Paul Devereux. “Recent developments in intergenerational mobility”. *Handbook of Labor Economics*. Ed. by Ashenfelter, Orley and David Card. Vol. 4.B. Amsterdam: North-Holland, 2011. Chap. 16, pp. 1487–1541.

[9] Blanden, Jo, Paul Gregg, and Lindsey Macmillan. “Accounting for intergenerational persistence”. *Economic Journal* 117 (2007), pp. C43–C60.

[10] Bloome, Deirdre. “Income inequality and intergenerational income mobility in the United States”. *Social Forces* 93.3 (2015), pp. 1047–1080.

[11] Bowles, Samuel and Herbert Gintis. “The inheritance of inequality”. *The Journal of Economic Perspectives* 16.3 (2002), pp. 3–30.

[12] Chadwick, Laura and Gary Solon. “Intergenerational income mobility among daughters”. *The American Economic Review* 92.1 (2002), pp. 335–344.

[13] Chernozhukov, Victor and Iván Fernández-Val. “Subsampling inference on quantile regression processes”. *Sankhyā: The Indian Journal of Statistics* (2005), pp. 253–276.

[14] Chernozhukov, Victor, Ivan Fernandez-Val, and Blaise Melly. “Inference on counterfactual distributions”. *Econometrica* 81.6 (2013), pp. 2205–2268.

[15] 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.

[16] 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.

[17] 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.

[18] Collins, William J and Gregory T Niemesh. "Unions and the great compression of wage inequality in the United States at mid-century". Working Paper. 2017.

[19] Collins, William J and Marianne H Wanamaker. "Up from slavery? African American intergenerational economic mobility since 1880". Working Paper. 2017.

[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] Donald, Stephen G, Yu-Chin Hsu, and Garry F Barrett. "Incorporating covariates in the measurement of welfare and inequality: methods and applications". *The Econometrics Journal* 15.1 (2012), pp. C1–C30.

[23] 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.

[24] Firpo, Sergio. "Efficient semiparametric estimation of quantile treatment effects". *Econometrica* 75.1 (2007), pp. 259–276.

[25] Firpo, Sergio, Nicole M Fortin, and Thomas Lemieux. "Unconditional quantile regressions". *Econometrica* 77.3 (2009), pp. 953–973.

[26] 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.

[27] Flores, Carlos A. "Estimation of dose-response functions and optimal doses with a continuous treatment". Working Paper. 2007.

[28] Flores, Carlos A, Alfonso Flores-Lagunes, Arturo Gonzalez, and Todd C Neumann. "Estimating the effects of length of exposure to instruction in a training program: the case of job corps". *Review of Economics and Statistics* 94.1 (2012), pp. 153–171.

[29] 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.

[30] 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.

[31] 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.

[32] Groves, Melissa Osborne. "Personality and the intergenerational transmission of economic status". *Unequal chances: Family background and economic success* (2005), pp. 208–231.

[33] Heckman, James, Hidehiko Ichimura, and Petra Todd. "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme". *The Review of Economic Studies* 64.4 (1997), pp. 605–654.

[34] 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.

[35] Hirano, Keisuke, Guido Imbens, and Geert Ridder. "Efficient estimation of average treatment effects using the estimated propensity score". *Econometrica* 71.4 (2003), pp. 1161–1189.

[36] Imbens, Guido W, Donald B Rubin, and Bruce I Sacerdote. "Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players". *American Economic Review* 91.4 (2001), pp. 778–794.

[37] Imbens, Guido and Jeffrey Wooldridge. "Recent developments in the econometrics of program evaluation". *Journal of Economic Literature* 47.1 (2009), pp. 5–86.

[38] Jantti, Markus, Bernt Bratsberg, Knut Roed, Oddbjorn Raaum, Robin Naylor, Eva Osterbacka, Anders Bjorklund, and Tor Eriksson. "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.

[39] Jasova, Martina, Caterina Mendicino, and Dominik Supera. "Rollover risk and bank lending behavior: Evidence from unconventional central bank liquidity". Working Paper. 2018.

[40] 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).

[41] Kitagawa, Toru, Martin Nybom, and Jan Stuhler. “Measurement error and rank correlations”. Working Paper. 2017.

[42] Koenker, Roger. *Quantile Regression*. Cambridge University Press, 2005.

[43] Koenker, Roger and Gilbert Bassett Jr. “Regression quantiles”. *Econometrica* (1978), pp. 33–50.

[44] Kosorok, Michael R. *Introduction to Empirical Processes and Semiparametric Inference*. Springer Science & Business Media, 2007.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] Melly, Blaise. “Decomposition of differences in distribution using quantile regression”. *Labour economics* 12.4 (2005), pp. 577–590.

[50] Melly, Blaise and Giulia Santangelo. “The changes-in-changes model with covariates”. Working Paper. 2015.

[51] Murtazashvili, Irina. “An alternative measure of intergenerational income mobility based on a random coefficient model”. *Journal of Applied Econometrics* 27.7 (2012), pp. 1161–1173.

[52] Murtazashvili, Irina, Di Liu, and Artem Prokhorov. “Two-sample nonparametric estimation of intergenerational income mobility in the United States and Sweden”. *Canadian Journal of Economics/Revue canadienne d’économique* 48.5 (2015), pp. 1733–1761.

[53] Ñopo, Hugo. “An extension of the Blinder–Oaxaca decomposition to a continuum of comparison groups”. *Economics Letters* 100.2 (2008), pp. 292–296.

[54] Richey, Jeremiah and Alicia Rosburg. “Decomposing economic mobility transition matrices”. Working Paper. 2015.

[55] Richey, Jeremiah and Alicia Rosburg. “Understanding intergenerational economic mobility by decomposing joint distributions”. Working Paper. 2016.

[56] Richey, Jeremiah and Alicia Rosburg. “Changing roles of ability and education in US intergenerational mobility”. *Economic Inquiry* 55.1 (2017), pp. 187–201.

[57] Rosenbaum, Paul and Donald Rubin. “The central role of the propensity score in observational studies for causal effects”. *Biometrika* 70.1 (1983), pp. 41–55.

[58] Rothe, Christoph. “Nonparametric estimation of distributional policy effects”. *Journal of Econometrics* 155.1 (2010), pp. 56–70.

[59] Solon, Gary. “Intergenerational income mobility in the United States”. *The American Economic Review* (1992), pp. 393–408.

[60] Solon, Gary. “Intergenerational mobility in the labor market”. *Handbook of labor economics* 3 (1999), pp. 1761–1800.

[61] Ulrick, Shawn W. “The Oaxaca decomposition generalized to a continuous group variable”. *Economics Letters* 115.1 (2012), pp. 35–37.

[62] Van Der Vaart, Aad W and Jon A Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, 1996.

[63] Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. MIT press, 2010.

[64] Wuthrich, Kaspar. “Semiparametric estimation of quantile treatment effects with endogeneity”. Working Paper. 2015.

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

(v) $E\|P_1(T,X)\|^{2+\varepsilon} < \infty$ *and* $E\|P_2(T)\|^{2+\varepsilon} < \infty$ *for some* $\varepsilon > 0$*.*

(vi) Let $J_1(\tau) = E\left[f_{Y|T,X}(P_1(T,X)'\alpha(\tau)|T,X)P_1(T,X)P_1(T,X)'\right]$*. Also, let*
$J_2(\tau) = E\left[f_{Y|T}(P_2(T)'\beta(\tau)|T)P_2(T)P_2(T)'\right]$*. The minimum eigenvalues of* $J_1(\tau)$ *and* $J_2(\tau)$ *are uniformly bounded away from zero.*

B Proofs

B.1 Proof of Theorem 1

Let

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

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 the outcome 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

$$\psi_1(W,\xi) = f_{Y|T,X}(y|t,x)P_1(t,x)'J_1(F_{Y|T,X}(y|t,x))^{-1}H_1(Y,T,X,F_{Y|T,X}(y|t,x))$$

where

$$H_1(Y,T,X,\tau) = \left(\mathbb{1}\{Y \leq P_1(T,X)'\alpha(\tau)\} - \tau\right) P_1(T,X)$$

Next, let

$$\psi_2(W,\xi) = f_{Y|T}(\bar{y}|\bar{t}) P_2(\bar{t})' J_2(F_{Y|T}(\bar{y}|\bar{t}))^{-1} H_2(Y,T,F_{Y|T}(y|t))$$

where

$$H_2(Y,T,\tau) = (\mathbb{1}\{Y \leq P_2(T)'\beta(\tau)\} - \tau))) \, P_2(T)$$

and let

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

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 $\mathbb{S} = l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{YT}) \times l^\infty(\mathcal{X})$. Under Assumptions 2 to 5 and Assumption A.1,*

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

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

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

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

Proof. The result follows immediately under Assumptions 5 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 $\psi : \mathbb{D}_\psi \subset \mathbb{D} = l^\infty(\mathcal{YTX}) \times l^\infty(\mathcal{X}) \mapsto l^\infty(\mathcal{YT})$ given by*

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

for $\Lambda = (\Lambda_1,\Lambda_2) \in \mathbb{D}$. Then, under Assumptions 2 to 5 and Assumption A.1, 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

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

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

$$\lambda_{1k} = \frac{\Lambda_{1k} - \Lambda_1}{t_k}$$
$$\lambda_{2k} = \frac{\Lambda_{2k} - \Lambda_2}{t_k}$$

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

Then,

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

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

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, under Assumptions 2 to 5 and Assumption A.1, 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*

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

Proof. The result follows immediately from Lemma 2. □

Proof of Theorem 1

Lemma 3 implies

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

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

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

($\mathbb{G}_{Y|T,X}$ 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 Tables and Figures

Table 1: Summary Statistics

	Q1	Q2	Q3	Q4	All
Parents' Income (1000s)	32.53	51.47	67.97	107.49	64.87
	(0.291)	(0.144)	(0.183)	(1.289)	(0.568)
Child's Income (1000s)	45.96	64.02	74.8	96.8	70.4
	(0.935)	(1.518)	(1.491)	(2.393)	(0.888)
Head White	0.72	0.93	0.95	0.96	0.89
	(0.015)	(0.008)	(0.007)	(0.006)	(0.005)
Head Non-White	0.28	0.07	0.05	0.04	0.11
	(0.015)	(0.008)	(0.007)	(0.006)	(0.005)
Child Male	0.49	0.47	0.48	0.5	0.48
	(0.017)	(0.017)	(0.017)	(0.017)	(0.008)
Head Male	0.77	0.95	0.97	0.99	0.92
	(0.014)	(0.007)	(0.006)	(0.003)	(0.004)
Year Born	1970.69	1970.29	1970.64	1969.32	1970.23
	(0.324)	(0.309)	(0.32)	(0.345)	(0.163)
Head Veteran	0.26	0.37	0.47	0.47	0.39
	(0.014)	(0.016)	(0.017)	(0.017)	(0.008)
Head Less than HS	0.35	0.19	0.08	0.05	0.17
	(0.016)	(0.013)	(0.009)	(0.007)	(0.006)
Head HS	0.58	0.64	0.58	0.39	0.55
	(0.016)	(0.016)	(0.016)	(0.016)	(0.008)
Head College	0.07	0.17	0.34	0.57	0.29
	(0.008)	(0.013)	(0.016)	(0.016)	(0.007)
Cutoff	44.24	59.17	78.01	434.44	
N	908	907	907	908	3630

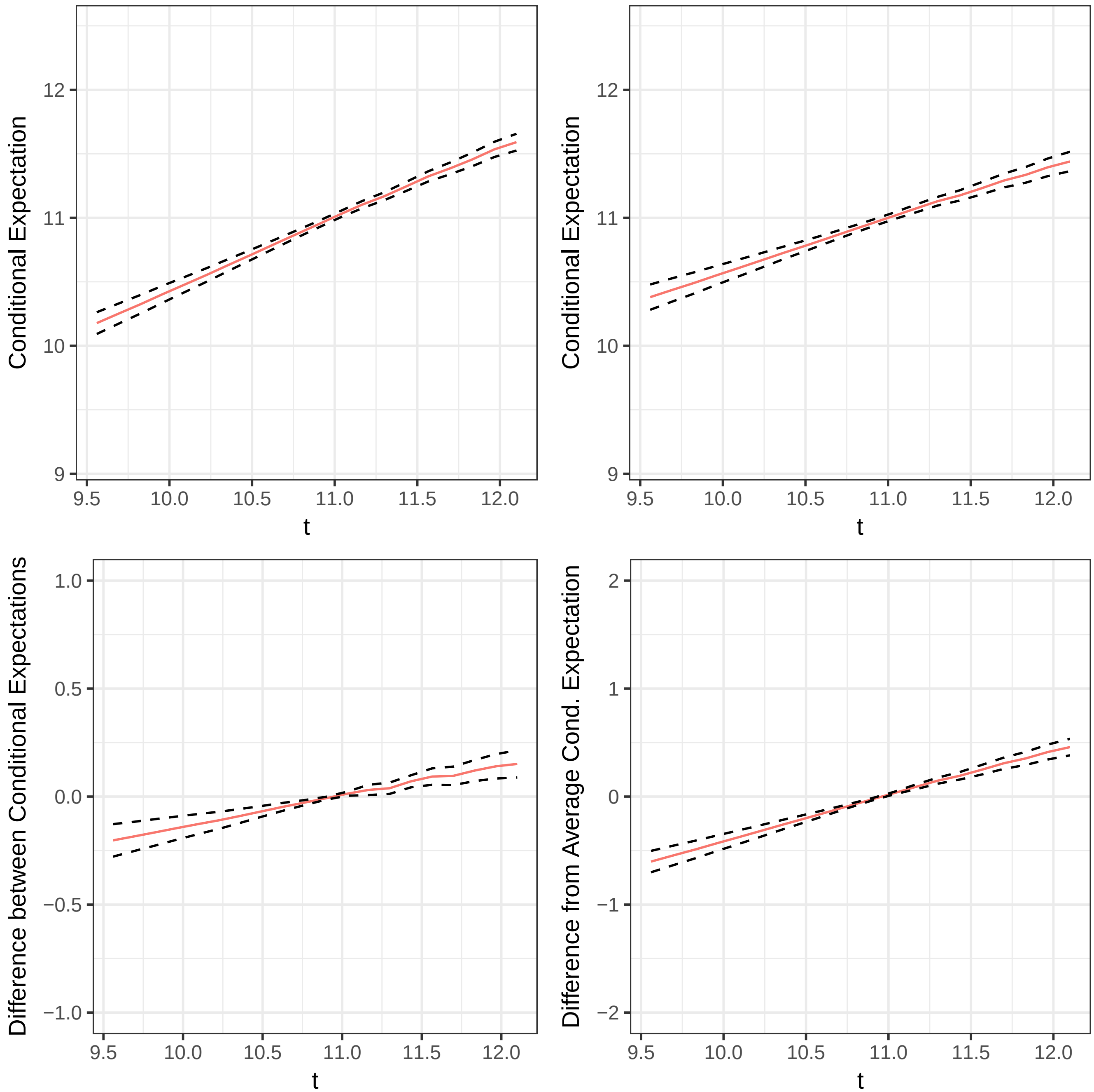
Notes: Summary statistics for the main dataset used in the paper. Each column provides average values of available variables by parents' income quartile. Standard errors are given in parentheses beneath the average. The row "Cutoff" is the maximum value of parents' income in that quartile (i.e. the dividing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

Table 2: Intergenerational Elasticity (IGE) Estimates

	Dependent variable:			
	Log Child's Income			
	(1)	(2)	(3)	(4)
Log Parents' Income	0.609***	0.573***	0.559***	0.452***
	(0.023)	(0.024)	(0.024)	(0.027)
Head Non-White		−0.245***	−0.253***	−0.236***
		(0.040)	(0.040)	(0.039)
Male		0.025	0.022	0.023
		(0.019)	(0.019)	(0.019)
Head Male		−0.076*	−0.056	−0.019
		(0.043)	(0.043)	(0.043)
Year Born			−0.009***	−0.013***
			(0.001)	(0.001)
Head Veteran				−0.006
				(0.021)
Head Less Than HS Educ.				−0.225***
				(0.030)
Head College Educ.				0.104***
				(0.024)
Constant	4.282***	4.761***	22.282***	31.110***
	(0.255)	(0.261)	(1.984)	(2.211)

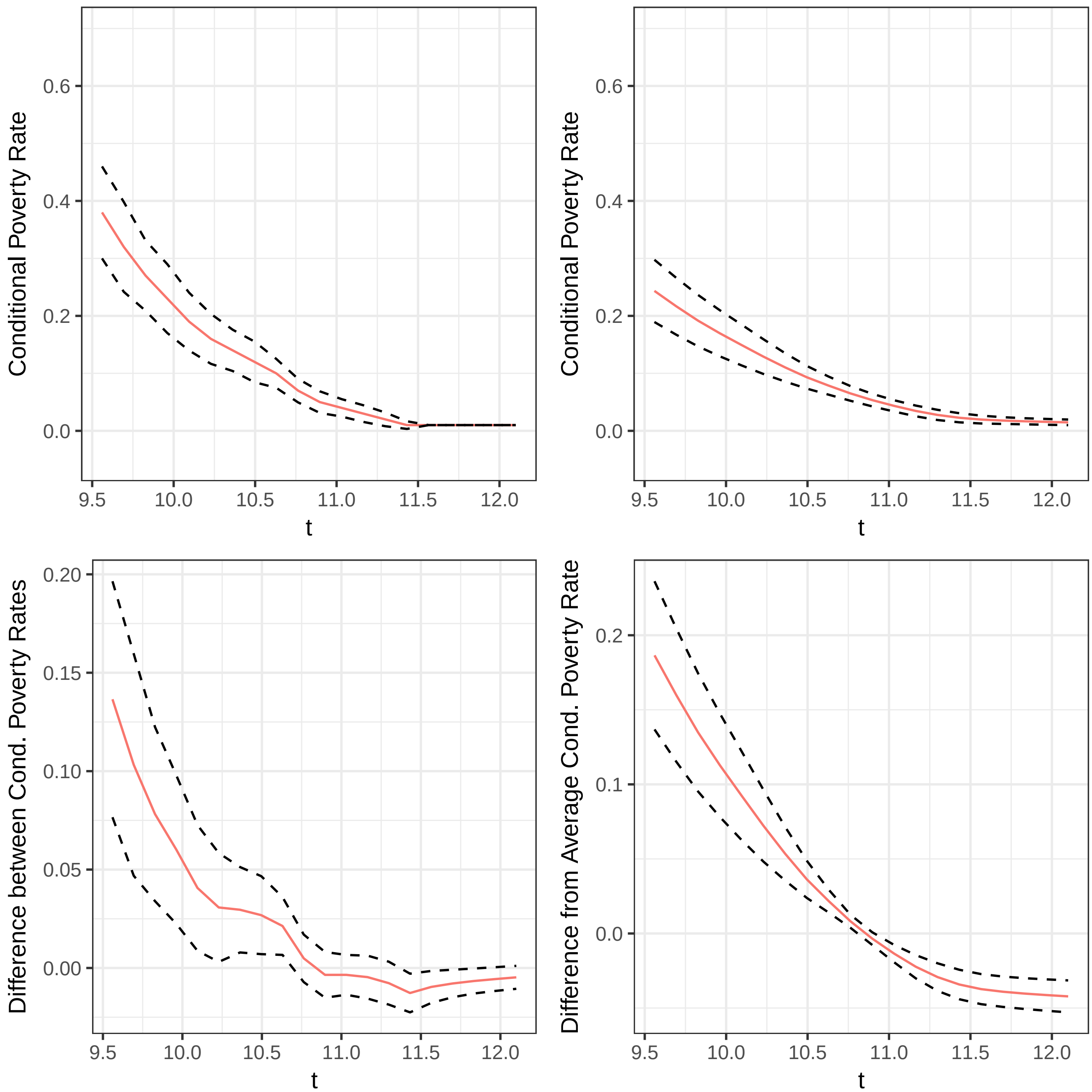
Notes: Results come from regressions of the log of child's income on the log of parents' income and additional controls using the full sample of 3,630 observations. Standard errors are heteroskedasticity robust.
Sources: Panel Study of Income Dynamics, as described in text

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



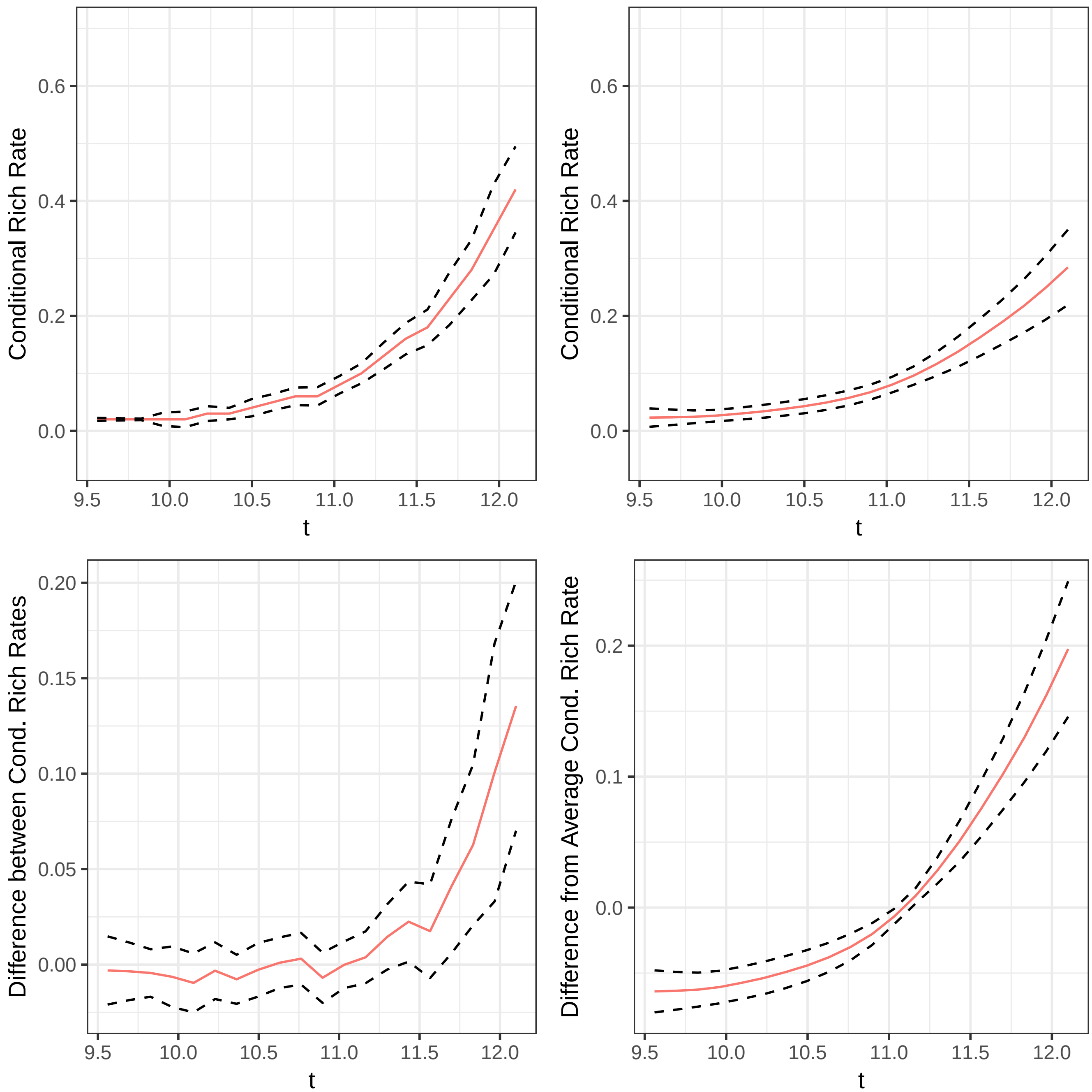
Notes: The top left panel plots average child's income as a function of parents' income with no adjustments for other covariates. The top right panel adjusts for differences in the covariates family head's race, family head's gender, gender of child, child's birth year, family head's veteran status, 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 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: Fraction of Children below the Poverty Line



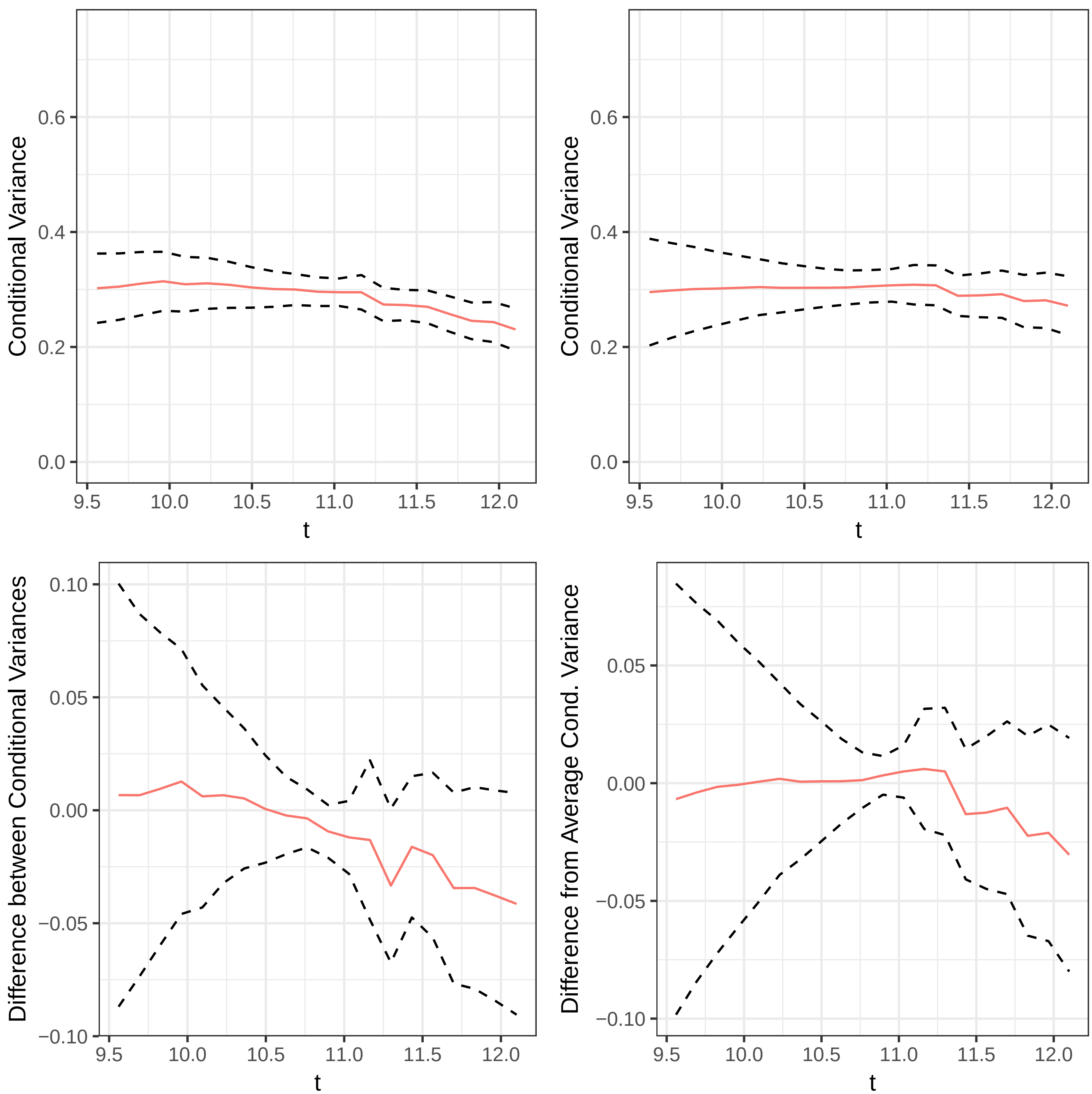
Notes: The top left panel plots the fraction of children below the poverty line as a function of parents' income with no adjustments for other covariates. The poverty line is set to be $22,113 which is the poverty line for a family with two adults and two children in 2010. The top right panel adjusts 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 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 "Rich" Children



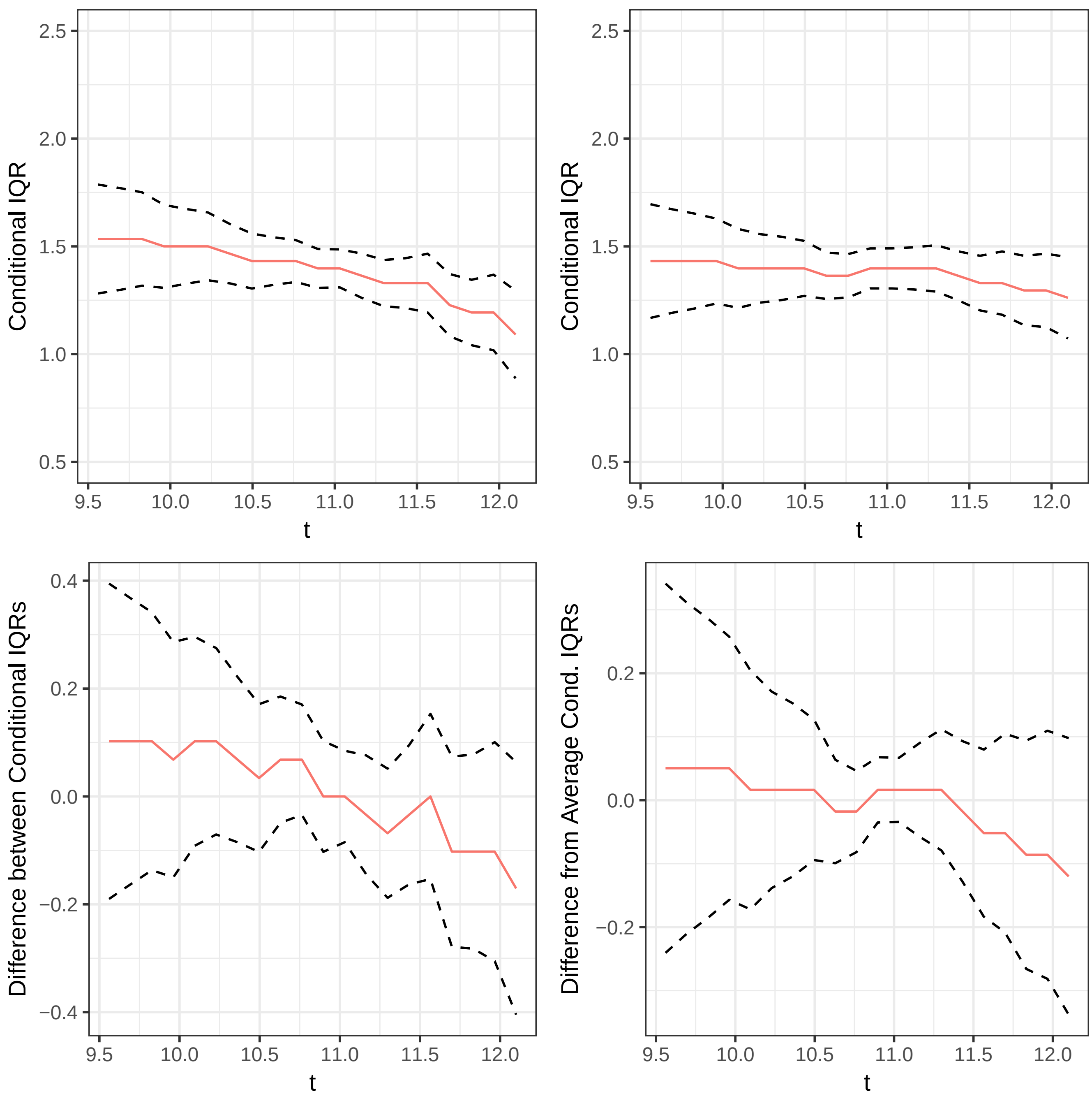
Notes: The top left panel plots the fraction of "rich" children as a function of parents' income with no adjustments for other covariates where "rich" is defined as having income above the 90th percentile of income in the U.S. in 2010 which is $132,923. The top right panel adjusts for differences in 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 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: 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 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: Inter-Quantile Ranges



Notes: The top left panel plots the inter-quantile 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