Semiparametric Estimation of Oaxaca-Blinder Decompositions with Continuous Groups

Brantly Callaway* Weige Huang†

October 8, 2018

Preliminary, please do not cite.

Abstract

This paper considers Oaxaca-Blinder type decompositions with continuous groups. In particular, we decompose the differences between outcomes at a series of values of the group variable and at a particular value of the group variable into (i) a composition effect and (ii) a structure effect. The composition effect is due to differences in the distribution of observed characteristics (e.g. race or education) for individuals at two particular values of the continuous group variable. The structure effect is due to difference in the "return" to characteristics at two particular values of the continuous group variable. We also consider detailed decompositions of both the composition and structure effects. Our procedure is based on semiparametric smooth varying coefficient models that are essentially "local" regressions in terms of the continuous group variable. This approach is distinct from previous work on decompositions with continuous groups that invoke parametric models. We develop the limiting distribution of our estimator and show the validity of the wild bootstrap for conducting inference. We apply our method to decompose earnings differentials for individuals across different values of their parents' income (i.e. parents' income is the continuous group).

Keywords: Oaxaca-Blinder Decomposition, Continuous Decomposition, Smooth Varying Coefficient Model, Local Regression, Semiparametric Methods, Intergenerational Income Mobility

JEL Codes: C14, J62

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

†Corresponding Author, PhD Student, Department of Economics, Temple University. Email: darren1988@163.com

1 Introduction

Since the seminal work of Oaxaca (1973) and Blinder (1973) in the early 1970s, there have been many studies decomposing differences in outcomes between two groups into parts that are explained by (i) differences in the distributions of observed characteristics between the two groups and (ii) "structural" differences between the two groups representing differences in the "return" to various characteristics between the two groups (often these are differences in estimated coefficients on observed characteristics between the two groups). To give some examples, Juhn, Murphy, and Pierce (1993) decompose wage inequality for males, Firpo, Fortin, and Lemieux (2009) consider the direct effect of union status on male log wage using unconditional quantile regression, and Chernozhukov, Fernandez-Val, and Melly (2013) decompose quantile differences using distribution regression.[1] Fortin, Lemieux, and Firpo (2011) offers a comprehensive overview of decomposition methods for differences between two groups.

Almost universally, these types of procedures are used to decompose the outcomes of two discrete groups, e.g., male and female. In this paper, we consider the case where the group variable is continuous (e.g., student test scores, years of work experience and parents' income) instead of binary or categorical (e.g., union status or races). There is a small literature on decompositions with a continuous group variable that includes Ñopo (2008) and Ulrick (2012). Both of these papers rely on linear models for the outcome in terms of the treatment and other covariates.[2] However, this approach is very limiting. To give an example, in our application on intergenerational income mobility, this setup (i) requires the effect of parents' income on child's income to be the same across different values of parents' income, and (ii) restricts how other covariates effect the outcome for different values of parents' income. These sorts of restrictions seem likely to be violated in many applications in economics, and we provide evidence that both of these conditions are likely to be violated in our application.

In this paper, we propose a decomposing differences in outcomes across different values of the

[1]In addition to these, Freeman (1980) looks at the difference in the variance of wages between the union and non-union sector. DiNardo, Fortin, and Lemieux (1996) analyze changes in the U.S. distribution of wages by applying kernel density methods. See also Donald, Green, and Paarsch (2000), Machado and Mata (2005), Altonji, Bharadwaj, and Lange (2012), Gelbach (2016), Elder, Goddeeris, and Haider (2015) among others.

[2]Relatedly, Richey and Rosburg (2016) and Callaway and Huang (2018) propose decompositions in the case with continuous groups that apply to the entire distribution, but these papers invoke distribution regression and quantile regression setups that put similar structure on the decomposition problem as in Ñopo (2008) and Ulrick (2012). The approach in the current paper is distinct from these.

group variable by embedding a first-step semiparametric smooth varying coefficient model (Li, Huang, Li, and Fu (2002))[3] into a traditional Oaxaca-Blinder decomposition. In particular, we propose estimating the expecation of the outcome conditional on the treatment and other covariates using a regression that is "local" in the continuous group variable and linear in the other variables. This approach offers many advantages relative to the fully parametric approach used in existing work. First, our approach allows the effect of the group variable on the outcome to vary (in arbitrary ways) across different values of the group variable. Second, our approach allows the effects of covariates to differ (in arbitrary ways) across different values of the group variable. For example, in our application on intergenerational income mobility, our approach allows the effect of race or parents' education to be different for children from low income families relative to high income families. An alternative idea would be to try to nonparametrically estimate the entire decomposition. Our approach offers several advantages relative to this idea. First, our semiparametric method does not suffer from the curse of dimensionality. In realistic applications with, for example, several thousand observations and 5-10 covariates, the nonparametric approach is likely to perform quite poorly. Second, we show that the decomposition results follow immediately using our approach (including detailed decompositions), but these would be quite difficult to obtain in the nonparametric estimation case.[4]

In addition, we are able to perform detailed decompositions that further decompose both the composition effect and the structure effect into components due to each of the observed covariates. Our decomposition results do not depend on the order of decomposition, that is, our results are path independent. We also establish the joint asymptotic normality of the decomposition components at a grid of values of the group variable and use the wild bootstrap to construct inference that covers particular components of the decomposition at all values in the grid of values for the continuous group variable with fixed probability.

We apply our method to decompose earnings differentials for children across their parents'

[3]Other work on smooth varying coefficient models includes Hastie and Tibshirani (1993), Cai, Fan, and Li (2000), Fan, Zhang, et al. (1999), Zhang, Lee, and Song (2002), and Li and Racine (2010) and among others.

[4]Another alternative idea would be to create a finite number of groups based on the value of the continuous group variable and then using ideas from the literature on decompositions with discrete groups (see, for example, Cameron and Heckman (2001), Frenette et al. (2007), Du, Renyu, He, and Zhang (2014) and Richey and Rosburg (2018))). This approach suffers from having to somehow choose where to create cutoffs for each group. Also, our "local" approach can be seen as a limiting case for this idea though we are able to circumvent the problem of arbitrarily choosing cutoffs.

income (i.e. parents' income is the continuous group) into parts due to differences in the distributions of education, race and other observable characteristics at different values of parents' income and into parts that are due to differences in the return to these characterics at different values of parents' income. We find that differences in outcomes are relatively more driven by the differences in the "return" to covariates (i.e. structure effects) than by the differences in the distribution of covariates (i.e. composition effects) across different values of parents' income. This is most pronounced for relatively high values of parents' income. We also find that the composition effects are mostly driven by parents' education and race for children from low income families; at higher values of parents' income, the effect of race is smaller. We also find that the return to parents' eduction is higher for children from low income families than for those from high income families.

We organize the rest of the paper as follows. Section 2 describes the parameters of interest. In Section 3, we present our new approach. The asymptotic theory of the decomposition components is developed in Section 4. Section 5 applies the method to decompose earnings differentials for children across their parents' income. We conclude in Section 6.

2 Parameters of Interest

Let Y denotes outcome variable, T a continuous group variable and X a $(1+\mathbb{K})\times 1$ vector of covariates including a constant. Our goal is to decompose differences in average between outcomes conditional on two particular values of a continuous group variable: $E(Y|T=t)-E(Y|T=t_0)$.[5] The difference can be written as:

$$\begin{aligned}
\Delta_O(t) &:= E(Y|T=t)-E(Y|T=t_0) \\
&= E\left(E(Y|X,T=t)|T=t\right)-E\left(E(Y|X,T=t_0|T=t_0)\right) \\
&= \left\{E\left(E(Y|X,T=t)|T=t\right)-E\left(E(Y|X,T=t)|T=t_0\right)\right\} \\
&\qquad\qquad + \left\{E\left(E(Y|X,T=t)|T=t_0\right)-E\left(E(Y|X,T=t_0|T=t_0)\right)\right\} \\
&= \int_{\mathcal{X}} E[Y|X=x,T=t]\left(dF_{X|T}(x|t)-dF_{X|T}(x|t_0)\right) \qquad (2.1)
\end{aligned}$$

[5] In the application, we found it most useful to hold t_0 constant (one can set it be equal to the median or some other value of the continuous group variable) and vary t across many different possible values of the continuous group variable. This strategy results in plots that are easy to interpret. All the arguments that we make below go through for each possible value of t; therefore, in this section we focus on a single value of t.

$$+ \int_{\mathcal{X}} \left(E[Y|X=x,T=t] - E[Y|X=x,T=t_0]\right) dF_{X|T}(x|t_0)$$

where the first equality is the definition of $\Delta_O(t)$, the second equality uses the law of iterated expectations, the third equality holds by adding and subtracting $E\left(E(Y|X,T=t)|T=t_0\right)$, and the last equality just re-expresses the previous line as an integral rather than an expectation. Equation (2.1) indicates that the overall difference between average outcomes for those with group value given by t relative to t_0 can be expressed as a difference due to differences in the distribution of covariates at each group value (this is called the composition effect and is what is given in the first term of Equation (2.1)) and another term due to differences in average outcomes for individuals who have the same value of the covariates but different values of the group variable (this is called the structure effect and is what is given in the second term of Equation (2.1)).

Notice that each of the terms in Equation (2.1) is identified by the sampling process. However, estimating each of these terms nonparametrically can be relatively challenging especially with medium size datasets; this is particularly true for the terms given by $E[Y|X,T]$ that are potentially high dimensional conditional expectations. The terms involving $F_{X|T}$ are somewhat easier as they only condition on the continuous group. The approach taken in existing work on decompositions with continuous groups is to specify a parametric model for $E[Y|X,T]$. This greatly simplifies the process of decomposing the differences in outcomes across groups, but it is subject to the strong possibility of misspecification.

Instead of imposing a linear model for $E[Y|X,T]$, we make the following assumption

Assumption 1 (Smooth Varying Coefficient Model)**.**

$$E[Y|X,T] = X'\beta(T)$$

Assumption 1 says that the expectation of the outcome conditional on the continuous group variable and other covariates is linear in X but that the coefficients can vary for different values of T. It allows for the effect of each element of X to change at different values of the continuous group variable T. Also, in our case, X includes a constant so that the effect of the group variable on the outcome can change across different values of the treatment. This setup is similar in spirit to including interactions between X and T, but our approach is more general. In the case where X only includes discrete covariates, if the model is fully saturated in the sense that it includes indicators for all possible values of X, then this approach is fully nonparametric. In general, though, this approach is semiparametric. Next, we provide simple expressions for the composition effect and structure effect given in Equation (2.1) under the smooth varying coefficient model.

Proposition 1 (Aggregate Composition and Structure Effects)**.** *Under Assumption 1, $\Delta_O(t)$, the difference between mean outcomes for group value t and group value t_0, is given by*

$$\begin{aligned}\Delta_O(t) &= \left(E(X'|T=t) - E(X'|T=t_0)\right)\beta(t) + E(X'|T=t_0)\left(\beta(t) - \beta(t_0)\right) \\ &:= \Delta_C(t) + \Delta_S(t)\end{aligned}$$

where $\Delta_C(t)$ represents the aggregate composition effect and $\Delta_S(t)$ represents the aggregate structure effect.

Proof. Starting with the first term in Equation (2.1), under Assumption 1, $E[Y|X=x, T=t] = x'\beta(t)$. This implies that $E[E[Y|X, T=t]|T=t] = E[X'|T=t]\beta(t)$ which corresponds to the first term in the Proposition. Similar arguments apply to each of the other three terms and imply the result. □

Next, we show that one can perform a detailed decomposition using our approach where the aggregate composition and structure effects are further decomposed into components corresponding to each individual characteristic.

Proposition 2 (Detailed Decompositions)**.** *Under Assumption 1, the composition effect can be decomposed as*

$$\begin{aligned}\Delta_C(t) &= \sum_{k=1}^{\mathbb{K}+1}\left(E(X_k|T=t) - E(X_k|T=t_0)\right)\beta_k(t) \\ &:= \sum_{k=1}^{\mathbb{K}+1}\Delta C_k(t)\end{aligned}$$

where each $\Delta_{C_k}(t)$ represents the contribution of the k-th covariate to the composition effect.[6]

In addition, under Assumption 1, the structure effect can be decomposed as

$$\Delta_S(t) = \sum_{k=1}^{\mathbb{K}+1} E(X_k|T=t_0)\left(\beta_k(t) - \beta_k(t_0)\right)$$

[6]Notice that $\Delta_{C_1}(t)$ (the composition effect for the intercept) is equal to zero by construction.

$$:= \sum_{k=1}^{\mathbb{K}+1} \Delta S_k(t)$$

where each $\Delta_{S_k}(t)$ represents the contribution of the k-th covariate to the structure effect.

Proof. The result follows immediately under Assumption 1 and from the result in Proposition 1 $\square$

Interestingly, unlike the aggregate decomposition, it is unclear how one could proceed to perform a detailed decomposition nonparametrically. Proposition 2 also indicates that our detailed decompositions are path independent.

Remark 1. *Our approach allows the heterogeneous response of the outcome to changes in distribution of covariates conditional on continuous group variable. That is, $\beta(t)$ indexed by t is a function of t allowing changes across the values of group variable. For instance, t is parents' income, a X is father's education, and Y is the children's income. Our model allows father's education to have different impacts on the children's income conditional on various parents' income. This encompasses the case in which t is binary.*

Remark 2. *The first parameters that we consider are $\Delta_C(t)$ and $\Delta_S(t)$ both indexed by group variable t, that is, both are functions of t and straightforward measures to plot. Our approach is useful because it allows one to examine the role that the variation of distribution of covariates plays across different values of t in terms of the difference between outcomes conditional on t. More specifically, we can test whether $(\Delta_C(t_1), \cdots, \Delta_C(t_q))$ is jointly statistically different from zero which includes testing the significance of a particular $\Delta_C(t)$ being different from zero as a special case. The estimate $\Delta_S(t)$ is also a function of t and straightforward to plot. This allows one to examine the role that the differences in the returns to covariates across different values of t play in term of differences between outcomes conditional on t.*

Remark 3. *To examine the role that the differences in the distributions of particular covariates plays in terms of difference between outcomes across different values of t, we consider the parameters $\Delta_{C_k}(t)$ which are straightforward to plot, as a function of t. For instance, we can examine the role that the distribution of parents' education plays in term of differences between child's incomes conditional on parents' income. We can plot the differences between the child's incomes which are determined by the difference in the distribution of parents' education against parents' incomes.*

To examine the role that the returns to particular covariates plays in terms of differences between outcomes conditional on t, we also consider the parameters $\Delta_{S_k}(t)$ which, as a function of t, are also straightforward to plot. For instance, we can examine the role that the return to parents' education plays in term of differences between child's incomes conditional on parents' incomes. We can plot the differences between child's incomes determined by the difference in the return to parents' education against parents' incomes.

3 Estimation

In this section we first describes semiparametric smooth coefficient models. Then, we use plug-in approach to compute the decomposition components shown in Proposition 1 and Proposition 2. Lastly, compute the decomposition components repeatedly for differences between outcomes at a series of values of and a fixed particular value of continuous group variable.

3.1 Semiparametric Smooth Coefficient Models

Semiparametric smooth coefficient models (Li, Huang, Li, and Fu (2002)) is

$$Y_i = X_i'\beta(T_i) + \epsilon_i \tag{3.1}$$

The local constant estimator (Li and Racine (2007), p.302) of the model above is given by the minimizer to the following optimization

$$\hat{\beta}(t) = \arg\min_{\beta(t)} \sum_{i=1}^{n} K_{h_t}(T_i - t)(Y_i - X_i'\beta(t))^2. \tag{3.2}$$

We rewrite this optimization as

$$\hat{\beta}(t) = \arg\min_{\beta(t)} \left\| W_{h_t}^{1/2}\big(Y - X\beta(t)\big) \right\|_2^2. \tag{3.3}$$

where Y is a $n \times 1$ vector, X is an $n \times (1 + \mathbb{K})$ matrix, $W_{h_t} = Diag(K(\frac{T_1 - t}{h_t}), \ldots, K(\frac{T_n - t}{h_t})$ and $\|\cdot\|_2$ denotes L-2 norm.

Thus, $\hat{\beta}(t)$ can be computed by

$$\hat{\beta}(t) = \left(X'W_{h_t}X\right)^{-1}X'W_{h_t}Y \tag{3.4}$$

In practice, we select a bandwidth which makes the bias terms in the decomposition components converge to zero.

3.2 Decomposition

In this paper, we use local linear regression to estimate $E(X|T)$. Then, combined with estimates $\hat{\beta}(\cdot)$, the components of decomposition for a difference are straightforward to compute by plugging $\hat{E}(X|T)$ and $\hat{\beta}(\cdot)$ into the results in Proposition 1 and Proposition 2. Then, one can repeatedly compute the decomposition components for differences between outcomes at a series of t's and a fixed particular t_0 using the same approach.

4 Asymptotic Theory

This section shows joint asymptotic normality of decomposition components in differences between outcomes at a series of values of and at a particular value of continuous variable, which including a difference between outcomes at two particular values of group variable as a special case. We establish theorems which allow one to conduct inferences on a series of decomposition components simultaneously, although the theorems are not for uniform inference.

For simplicity, we assume all bandwidths converge at the same rate as h. That is, $h_t = C_t h$ and $h_k = C_k h, k = 1, \cdots, \mathbb{K}$ where h_k is the bandwidth used in estimating $E(X_k|T = \cdot)$ based on local linear regression, and C_t and C_k are both some constants.

Assumption 2. *Assume all bandwidths converge at the same rate as h: $h_t = C_t h$ and $h_k = C_k h$, where C_t and C_k are some constants.*

Let $X = g(T) + \epsilon$. We also assume the regularity conditions for local linear regression model (see, for example, Li and Racine (2007)[Theorem 2.7]) and local constant estimator of semiparametric smooth coefficient model (Li, Huang, Li, and Fu (2002)).

4.1 Joint Asymptotic Normality of Decomposition Components

Four theorems below establish respectively the asymptotic normality of the estimated decomposition components. κ_2 and $g''(t)$ shown in the following theorems are defines as $\kappa_2 = \int k(v)v^2 dv$, $g''(t) = \partial^2 g(t)/\partial t^2$.

4.1.1 Aggregate Components

Theorem 1 establishes the joint asymptotic normality of the aggregate composition affects across a series of values of the group variable from t_1 to t_q.

Theorem 1. *Under regularity conditions stated in Lemma 1 and 2 and Assumption 2, the joint asymptotic normality of the aggregate composition effect $\hat{\Delta}_C(t)$'s across a series of values of the group variable from t_1 to t_q is as follows*

$$\sqrt{nh}\left(\hat{\Delta}_C(t_1) - \Delta_C(t_1) - Bias_{\Delta_C}(t_1), \quad \cdots, \quad \hat{\Delta}_C(t_q) - \Delta_C(t_q) - Bias_{\Delta_C}(t_q)\right)' \xrightarrow{d} N(0, \mathbb{V}_C)$$

where $\mathbb{V}_C = R_{\Delta_C}\Sigma_{\Delta_C}R'_{\Delta_C}$ *where* R_{Δ_C}, Σ_{Δ_C} *and* $Bias_{\Delta_C}(\cdot)$ *are given in section A.1 of the appendix.*

Theorem 1 shows the joint asymptotic normality of the aggregate composition effects of differences between outcomes conditional on a series of values and a particular reference value of a group variable which includes as a special case one aggregate composition effect of difference between outcomes conditional on one value and a reference value of group variable. Although Theorem 1 does not enable uniform inference, it is useful because it allows one to consider inference on a list of aggregate composition effects and will be important for testing whether the aggregate composition effects are jointly significantly different from zero.

Simultaneous consistency of $\Delta_C(t_1)$ through $\Delta_C(t_q)$ requires that $nh \to 0$ and $h^2 \to 0$ in order for the bias and the variances to both converge to zero.

Theorem 2 establishes the joint asymptotic normality of the aggregate structure effects of differences between outcomes conditional on a series of values and a particular reference value of a group variable which includes as a special case one aggregate composition effect of difference between outcomes conditional on one value and a reference value of group variable.

Theorem 2. *Under regularity conditions stated in Lemma 1 and 2 and Assumption 2, the joint asymptotic normality of the aggregate composition effect $\hat{\Delta}_S(t)$'s across a series of values of the group variable from t_1 to t_q is as follows*

$$\sqrt{nh}\left(\hat{\Delta}_S(t_1) - \Delta_S(t_1) - Bias_{\Delta_S}(t_1), \quad \cdots, \quad \hat{\Delta}_S(t_q) - \Delta_S(t_q) - Bias_{\Delta_S}(t_q)\right)' \xrightarrow{d} N(0, \mathbb{V}_S)$$

where $\mathbb{V}_S = R_{\Delta_S}\Sigma_{\Delta_S}R'_{\Delta_S}$ *where* R_{Δ_S}, Σ_{Δ_S} *and* $Bias_{\Delta_S}(\cdot)$ *are given in section A.2 of the appendix.*

Although Theorem 2 does not allow uniform inference, it is useful because it allows one to consider inference on a list of aggregate composition effects thus will be important for testing whether the aggregate structure effects are jointly significantly different from zero. Simultaneous consistency of $\Delta_S(t_1)$ through $\Delta_S(t_q)$ requires that $nh \to 0$ and $h^2 \to 0$ in order for the bias and the variances to both converge to zero.

4.1.2 Detailed Components

The two following theorems establish respectively the joint asymptotic normality of the detailed composition and structure effects of . Theorem 3 shows the joint asymptotic normality of the detailed composition effects of differences between outcomes conditional on a series of values and a particular reference value of a group variable which includes as a special case one detailed composition effect.

Theorem 3. *Under regularity conditions stated in Lemma 1 and 2 and Assumption 2, the joint asymptotic normality of the detailed composition effect $\hat{\Delta}_{C_k}(t)$'s across a series of values of the group variable from t_1 to t_q is as follows*

$$\sqrt{nh}\left(\hat{\Delta}_{C_k}(t_1) - \Delta_{C_k}(t_1) - Bias_{\Delta_{C_k}}(t_1), \quad \cdots, \quad \hat{\Delta}_{C_k}(t_q) - \Delta_{C_k}(t_q) - Bias_{\Delta_{C_k}}(t_q)\right)'$$
$$\xrightarrow{d} N(0, \mathbb{V}_{C_k})$$

where $\mathbb{V}_{C_k} = R_{\Delta_{C_k}}\Sigma_{\Delta_{C_k}}R'_{\Delta_{C_k}}$ *where* $R_{\Delta_{C_k}}$, $\Sigma_{\Delta_{C_k}}$ *and* $\Delta_{C_k}(\cdot)$ *are given in section A.3 of the appendix.*

Theorem 3 allows one to jointly test whether the difference in distribution of a covariate conditional on the group variable contribute significantly to the differences between outcomes conditional a series of values and a particular value of group variable. Simultaneous consistency of $\Delta_{C_k}(t_1)$ through $\Delta_{C_k}(t_q)$ requires that $nh \to 0$ and $h^2 \to 0$ in order for the bias and the variances to both converge to zero.

Theorem 4 establishes the joint asymptotic normality of the detailed structure effects. It is useful because it allows one to jointly test whether a return to a covariate significantly contributes to the differences between outcomes conditional a series of values and a particular value of group variable.

Theorem 4. *Under regularity conditions stated in Lemma 1 and 2 and Assumption 2, the joint asymptotic normality of the detailed structure effect $\hat{\Delta}_{S_k}(t)$'s across a series of values of the group variable from t_1 to t_q is as follows*

$$\sqrt{nh}\left(\hat{\Delta}_{S_k}(t_1) - \Delta_{S_k}(t_1) - Bias_{\Delta_{S_k}}(t_1), \quad \cdots, \quad \hat{\Delta}_{S_k}(t_q) - \Delta_{S_k}(t_q) - Bias_{\Delta_{S_k}}(t_q)\right)'$$
$$\xrightarrow{d} N(0, \mathbb{V}_{S_k})$$

where $\mathbb{V}_{S_k} = R_{\Delta_{S_k}}\Sigma_{\Delta_{S_k}}R'_{\Delta_{S_k}}$ *where* $R_{\Delta_{S_k}}$, $\Sigma_{\Delta_{S_k}}$ *and* $\Delta_{S_k}(\cdot)$ *are given in section A.4 of the appendix.*

Simultaneous consistency of $\Delta_{S_k}(t_1)$ through $\Delta_{S_k}(t_q)$ requires that $nh \to 0$ and $h^2 \to 0$ in order for the bias and the variances to both converge to zero.

In practice, we carry out inference using bootstrap.

5 Application on Intergenerational Income Mobility

Much research has documented that parents' income is correlated with the income of their children (see Solon (1992) and Solon (1999), among many others). However, there are many other "covariates" that are also correlated with both parents' income and child's income– for example, education and race. We apply new decomposition approaches stated above to decompose differences between child's incomes at a series of values of and the mean of parents' incomes into other observable characteristics. The decomposition allows us to examine the roles that these characteristics play in explaining the differences at various parents' income levels. We suspect the roles that family head's education and race contribute to the differences would vary across parents' incomes because the distribution of and/or the "returns" to these two characteristics conditional on parents' income are likely various with parents' income. That is, low-income families are likely having less educated family head and high-income families are likely to have family head with higher education. Also, the returns to the same education level could be not the same among families with different family incomes. Although both family heads have the same education levels, one family head, who has higher income, could affect the child – and therefore the child's income – differently from the other family heads, who have lower income.

For any decomposition dealing with dummy variables, we face the omitted variable problem.[7] In this paper, we choose to omit the following variables: white, female child, non-veteran head and head with high school education.[8] Also, the group chosen as reference could also affect the decomposition results. We show the main results using as reference the coefficients from smooth coefficient model in which the child's income at a series of values of parents' income is the dependent variable. We show the results as robustness checks by switching the reference. The advantage of our approach is that the order of decomposition does not matter, that is, the decomposition results are path independent.

To perform the decomposition, the first step is to choose the bandwidth for the smooth coefficient model. The bandwidth chosen needs to make the bias terms converge to zero. In practice, we consider bandwidths of the form $h = Cn^{\delta}$ for some values of C and δ; when $\delta = -1/5$, one can choose the C so as to obtain the "optimal" bandwidth using cross-validation. Our asymptotic results, however, require that we "undersmooth." Thus, ultimately, the bandwidth that we choose is given by $h = Cn^{-1/6}$ where C can be calculated using cross-validation. We use the same bandwidth for local linear regression of X on T.[9]

5.1 Data Description

The dataset used in this paper is from Callaway and Huang (2018). It has been a long history for researchers using Panel Study of Income Dynamics (PSID) to study intergenerational income mobility. In the paper, the outcome variable is children's family income and the continuous group

[7] There exists the "omitted group" problem in the decomposition and there is no general solution to this problem yet. See Jones (1983), Oaxaca and Ransom (1999) and Gelbach (2016).

[8] For the sake of comparison with the literature, we choose to omit high school graduates. See (Fortin, Lemieux, and Firpo (2011) p. 42), " ...the common practice of using high school graduates as the omitted category allows the comparison of detailed decomposition results when this omitted category is comparable across studies."

[9] In choosing the bandwidth using cross-validation, we generally found that the optimal bandwidth becomes smaller as fewer variables are included.

variable is parents' family income. The covariates include child's gender and year of birth, family head's race, educational attainment, and veteran status.

5.2 Results

Figure 1 shows the decomposition results. For the aggregate decomposition, we show that the aggregate components change across the parents' incomes. On average, the ratio of the composition effect to the overall estimated difference is higher when the parents' income is in the lower percentiles; the ratio is lower when the parents' income is in the higher percentiles. For example, at the 10th percentile of parents' incomes, differing characteristics between children from different families explain 40% of the estimated difference, with differing returns for these characteristics explaining the remaining difference. However, at the 90th percentile of parents' incomes, differing characteristics between children from different families explain 24.5% of the estimated difference, with differing returns for these characteristics explaining the remaining gap. The average ratio of the composition effect to the overall estimated difference is about 36.7% when the parents' income is within 20th and 40th percentiles. However, that average ratio decreases to 24.9% when the parents' income is within 60th and 90th percentiles.

Panel b of Figure 1 shows that the differing education explains most of the composition effect. Another interesting finding is that the differences in the distribution of race plays a significant role when the parents' income is in the lower percentiles; however, the differences in race distribution does not seem to contribute to the composition effect when the parents' income is in the higher percentiles.

Panel c of Figure 1 shows that the structure effect is most contributed by the returns for the other characteristics and the difference between the constant terms. We note an important finding that is the return to the education positively and significantly contributes to the structure effect when the parents' income is in the lower percentiles. This means that the return to education for the children from lower income families is higher than the one for the children from families with average income. This finding has important policy implications. It is also surprising to find that the return to race does not change across parents' incomes.

Figure 2 to 3 plots the overall differences and the effects across parents' incomes with %95 confidence bands obtained by bootstrapping. Figure 5 to 8 in the Appendix B show the decomposition results in which we change the reference group used to compute the effects.

6 Conclusion

The paper has developed a semiparametric method to decompose differences between outcomes conditional on different values of a continuous group variable, which can be useful in many interesting applications. We apply this method to study the effects of characteristics of family background like race and parents' education attainments on the differences between child's income conditional on parents' income. We find that conditional on parents' income the differences between child's income are mostly from the differences in the "returns" to family characteristics. We also find that the "return" to parents' education attainments is higher for children from low income family and the "return" to race is not significant different across different values of parents' income. Also, we find that the composition effects across parents' income are mostly driven by parents' education, especially from children from low income families.

References

[1] Altonji, Joseph G, Prashant Bharadwaj, and Fabian Lange. "Changes in the characteristics of American youth: Implications for adult outcomes". *Journal of Labor Economics* 30.4 (2012), pp. 783–828.

[2] Blinder, Alan S. "Wage discrimination: reduced form and structural estimates". *Journal of Human resources* (1973), pp. 436–455.

[3] Cai, Zongwu, Jianqing Fan, and Runze Li. "Efficient estimation and inferences for varying-coefficient models". *Journal of the American Statistical Association* 95.451 (2000), pp. 888–902.

[4] Callaway, Brantly and Weige Huang. "Distributional effects of a continuous treatment with an application on intergenerational mobility". Working Paper. 2018.

[5] Cameron, Stephen V and James J Heckman. "The dynamics of educational attainment for black, hispanic, and white males". *Journal of political Economy* 109.3 (2001), pp. 455–499.

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

[7] 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. ISSN: 00129682, 14680262. URL: `http://www.jstor.org/stable/2171954`.

[8] Donald, Stephen G, David A Green, and Harry J Paarsch. "Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates". *The Review of economic studies* 67.4 (2000), pp. 609–633.

[9] Du, Zaichao, LI Renyu, Qinying He, and Lin Zhang. "Decomposing the rich dad effect on income inequality using instrumental variable quantile regression". *China Economic Review* 31 (2014), pp. 379–391.

[10] Elder, Todd E, John H Goddeeris, and Steven J Haider. "Isolating the roles of individual covariates in reweighting estimation". *Journal of Applied Econometrics* 30.7 (2015), pp. 1169–1191.

[11] Fan, Jianqing, Wenyang Zhang, et al. “Statistical estimation in varying coefficient models”. *The annals of Statistics* 27.5 (1999), pp. 1491–1518.

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

[13] Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. “Decomposition methods in economics”. *Handbook of labor economics* 4 (2011), pp. 1–102.

[14] Freeman, Richard B. “Unionism and the Dispersion of Wages”. *ILR Review* 34.1 (1980), pp. 3–23.

[15] Frenette, Marc et al. *Why Are Youth from Lower-Income Families Less Likely to Attend University?: Evidence from Academic Abilities, Parental Influences, and Financial Constraints.* Statistics Canada Ottawa, 2007.

[16] Gelbach, Jonah B. “When do covariates matter? And which ones, and how much?” *Journal of Labor Economics* 34.2 (2016), pp. 509–543.

[17] Hastie, Trevor and Robert Tibshirani. “Varying-coefficient models”. *Journal of the Royal Statistical Society. Series B (Methodological)* (1993), pp. 757–796.

[18] Jones, Frank Lancaster. “On decomposing the wage gap: a critical comment on Blinder’s method”. *The Journal of Human Resources* 18.1 (1983), pp. 126–130.

[19] Juhn, Chinhui, Kevin M Murphy, and Brooks Pierce. “Wage inequality and the rise in returns to skill”. *Journal of political Economy* 101.3 (1993), pp. 410–442.

[20] Li, Qi, Cliff J Huang, Dong Li, and Tsu-Tan Fu. “Semiparametric smooth coefficient models”. *Journal of Business & Economic Statistics* 20.3 (2002), pp. 412–422.

[21] Li, Qi and Jeffrey S Racine. “Smooth varying-coefficient estimation and inference for qualitative and quantitative data”. *Econometric Theory* 26.6 (2010), pp. 1607–1637.

[22] Li, Qi and Jeffrey Scott Racine. *Nonparametric econometrics: theory and practice*. Princeton University Press, 2007.

[23] 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.

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

[25] Oaxaca, Ronald. "Male-female wage differentials in urban labor markets". *International economic review* (1973), pp. 693–709.

[26] Oaxaca, Ronald L and Michael R Ransom. "Identification in detailed wage decompositions". *The Review of Economics and Statistics* 81.1 (1999), pp. 154–157.

[27] Richey, Jeremiah and Alicia Rosburg. "Decomposing Joint Distributions via Reweighting Functions: An Application to Intergenerational Economic Mobility" (2016).

[28] Richey, Jeremiah and Alicia Rosburg. "Decomposing economic mobility transition matrices". *Journal of Applied Econometrics* 33.1 (2018). jae.2578, pp. 91–108. ISSN: 1099-1255. URL: http://dx.doi.org/10.1002/jae.2578.

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

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

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

[32] Zhang, Wenyang, Sik-Yum Lee, and Xinyuan Song. "Local polynomial fitting in semivarying coefficient model". *Journal of Multivariate Analysis* 82.1 (2002), pp. 166–188.

A Proofs

Before we show the proof of Theorem 1 to 4, we first prove the following three Lemmas. Lemma 1 show some results about the estimate from semiparametric smooth coefficient model; Lemma 2 shows results from local linear regression. Lemma 3 shows results on the components in $\hat{\Delta}_C(t)$.

We assume regularity conditions (Li, Huang, Li, and Fu (2002)) and other assumptions required for Theorem 9.3 (Li and Racine (2007)).

Lemma 1. *Under some regular conditions (see Li, Huang, Li, and Fu (2002)), one can have*

$$\begin{aligned}\hat{\beta}(t) - \beta(t) - Bias_t &= (nh_t)^{-1}\sum_{i=1}^{n} H_t^{-1} X_i \epsilon_i K(\frac{T_i - t}{h_t}) + (s.o.) \\ &:= (nh_t)^{-1}\sum_{i=1}^{n} \psi_1(t) + (s.o.)\end{aligned}$$

where

$$\begin{aligned}Bias_t &= \kappa_2 h_t^2 E[XX'\{\beta'(t) f_t'(X,t)/f(X|t) + \frac{1}{2}\beta''(t) f(t)\}|t] \\ \psi_1(t) &= H_t^{-1} X_i \epsilon_i K(\frac{T_i - t}{h_t}) \\ H_t &= f(t) E(XX'|T = t)\end{aligned}$$

Proof. See the proof of Theorem 9.3 in Li and Racine (2007). □

Lemma 2. *Let $X = g(T) + \epsilon$ and under standard regularity conditions for local linear estimators (see, for example, Li and Racine (2007)[Theorem 2.7]), one can show that*

$$\begin{aligned}\hat{E}(X|T = t) - E(X|T = t) - Bias_X(t) &= (nh_X)^{-1}\sum_{i=1}^{n} \frac{1}{f(t)} K(\frac{T_i - t}{h_X}) U_i + (s.o.) \\ &:= (nh_X)^{-1}\sum_{i=1}^{n} \psi_2(t) + (s.o.)\end{aligned}$$

where $Bias_X(t) = \frac{\kappa_2}{2} f(t) g_X''(t) h_X^2$ and $\psi_2(t) = \frac{1}{f(t)} K(\frac{T_i - t}{h_X}) U_i$.

Proof. The proof follows from Li and Racine (2007) Theorem 2.7. □

A.1 Proof of Theorem 1

To show the joint asymptotic normality of the aggregate composition effects stated in Theorem 1, we first introduce the next lemma which is a result about $\hat{\Delta}_C - \Delta_C$.

Lemma 3. *Under regularity conditions stated in Lemma 1 and 2,*

$$\hat{\Delta}_C(t) - \Delta_C(t) = Bias_{\Delta_C}(t) + R_{\Delta_C}(t)W_{\Delta_C}(t) + (s.o.)$$

where

$$\begin{aligned}
Bias_{\Delta_C}(t) &= \left(Bias_X(t) - Bias_X(t_0)\right)\beta(t) + \left(E(X'|T=t) - E(X'|T=t_0)\right) Bias_t \\
R_{\Delta_C}(t) &= \Big(\beta(t)', \quad -\beta(t)', \quad \big(E(X'|T=t) - E(X'|T=t_0)\big)\Big), \\
W_{\Delta_C}(t) &= \left(\left((nh_X)^{-1}\sum_{i=1}^n \psi_2(t)\right)', \left((nh_X)^{-1}\sum_{i=1}^n \psi_2(t_0)\right)', \left((nh_t)^{-1}\sum_{i=1}^n \psi_1(t)\right)'\right)'
\end{aligned}$$

.

Proof. The proof of Lemma 3 mainly follows three steps.

(i) Subtract $\hat{E}(X'|T=t)$ with $E(X'|T=t)$, $\hat{E}(X'|T=t_0)$ with $E(X'|T=t_0)$, and $\hat{\beta}(t)$ with $\beta(t)$ while adding or subtracting related terms to keep the equality hold.

(ii) Rearrange to have the terms stated in Lemma 1 and 2.

(iii) Apply the Lemma 1 and 2 to complete the proof.

$$\begin{aligned}
&\hat{\Delta}_C(t) - \Delta_C(t) \\
&= \left(\hat{E}(X'|T=t) - \hat{E}(X'|T=t_0)\right)\hat{\beta}(t) - \left(E(X'|T=t) - E(X'|T=t_0)\right)\beta(t) \\
&\qquad\qquad\qquad \text{Follow step (i)} \\
&= \underbrace{\left(\left(\hat{E}(X'|T=t) - E(X'|T=t)\right) - \left(\hat{E}(X'|T=t_0) - E(X'|T=t_0)\right)\right)\left(\hat{\beta}(t) - \beta(t)\right)}_{A} \\
&\quad + \hat{E}(X'|T=t)\beta(t) + E(X'|T=t)\left(\hat{\beta}(t) - \beta(t)\right) - \hat{E}(X'|T=t_0)\beta(t) - E(X'|T=t_0)\left(\hat{\beta}(t) - \beta(t)\right) \\
&\quad - \left(E(X'|T=t) - E(X'|T=t_0)\right)\beta(t) \\
&\qquad\qquad\qquad \text{Follow step (ii)}
\end{aligned}$$

$$
\begin{aligned}
&= \left(\hat{E}(X'|T=t) - E(X'|T=t)\right)\beta(t) - \left(\hat{E}(X'|T=t_0) - E(X'|T=t_0)\right)\beta(t) \\
&\quad + \left(E(X'|T=t) - E(X'|T=t_0)\right)\left(\hat{\beta}(t) - \beta(t)\right) \\
&\qquad\qquad\qquad\qquad \text{Follow step (iii)} \\
&= \left(Bias_X(t) + (nh_X)^{-1}\sum_{i=1}^{n}\psi_2(t)\right)\beta(t) - \left(Bias_X(t_0) + (nh_X)^{-1}\sum_{i=1}^{n}\psi_2(t_0)\right)\beta(t) \\
&\quad + \left(E(X'|T=t) - E(X'|T=t_0)\right)\left(Bias_t + (nh_t)^{-1}\sum_{i=1}^{n}\psi_1(t)\right) + (s.o.) \\
&:= Bias_{\Delta_C}(t) + R_{\Delta_C}(t)W_{\Delta_C}(t) + (s.o.)
\end{aligned}
$$

where it is straightforward to show the 3th equality holds using standard arguments that Term A is asymptotically negligible.

$\square$

Proof of Theorem 1.

Following Lemma 3, it is straightforward to prove Theorem 1 by Liapunov Central Limit Theorem.

$$
\begin{aligned}
&\left(\hat{\Delta}_C(t_1) - \Delta_C(t_1) - Bias_{\Delta_C}(t_1), \quad \cdots, \quad \hat{\Delta}_C(t_q) - \Delta_C(t_q) - Bias_{\Delta_C}(t_q)\right)' \\
&= \left(R_{\Delta_C}(t_1)W_{\Delta_C}(t_1), \quad \cdots, \quad R_{\Delta_C}(t_q)W_{\Delta_C}(t_q)\right)' + (s.o.) \\
&= R_{\Delta_C}W_{\Delta_C} + (s.o.)
\end{aligned}
$$

where

$$
\begin{aligned}
R_{\Delta_C} &= Diag\left(R_{\Delta_C}(t_1), \quad \cdots, \quad R_{\Delta_C}(t_q)\right), \\
W_{\Delta_C} &= \left(W'_{\Delta_C}(t_1), \quad \cdots, \quad W'_{\Delta_C}(t_q)\right)'
\end{aligned}
$$

and the first equality holds by Lemma 3.

Under Assumption 2, the proof is completed by Liapunov Central Limit Theorem.

$$
\sqrt{nh}\left(\hat{\Delta}_C(t_1) - \Delta_C(t_1) - Bias_{\Delta_C}(t_1), \quad \cdots, \quad \hat{\Delta}_C(t_q) - \Delta_C(t_q) - Bias_{\Delta_C}(t_q)\right)' \xrightarrow{d} N(0, \mathbb{V}_C)
$$

where

$$
\begin{aligned}
\mathbb{V}_C &= R_{\Delta_C}\Sigma_{\Delta_C}R'_{\Delta_C},\\
\Sigma_{\Delta_C} &= E\left(nhW_{\Delta_C}W'_{\Delta_C}\right)\\
&= nhE\begin{pmatrix}
W_{\Delta_C}(t_1)W'_{\Delta_C}(t_1) & W_{\Delta_C}(t_1)W'_{\Delta_C}(t_2) & \cdots & W_{\Delta_C}(t_1)W'_{\Delta_C}(t_q)\\
W_{\Delta_C}(t_2)W'_{\Delta_C}(t_1) & W_{\Delta_C}(t_2)W'_{\Delta_C}(t_2) & \cdots & W_{\Delta_C}(t_2)W'_{\Delta_C}(t_q)\\
\vdots & \vdots & \ddots & \vdots\\
W_{\Delta_C}(t_q)W'_{\Delta_C}(t_1) & W_{\Delta_C}(t_q)W'_{\Delta_C}(t_2) & \cdots & W_{\Delta_C}(t_q)W'_{\Delta_C}(t_q)
\end{pmatrix}\\
&= \begin{pmatrix}
\Sigma^{11}_{\Delta_C} & \Sigma^{12}_{\Delta_C} & \cdots & \Sigma^{1q}_{\Delta_C}\\
\Sigma^{21}_{\Delta_C} & \Sigma^{22}_{\Delta_C} & \cdots & \Sigma^{2q}_{\Delta_C}\\
\vdots & \vdots & \ddots & \vdots\\
\Sigma^{q1}_{\Delta_C} & \Sigma^{q2}_{\Delta_C} & \cdots & \Sigma^{qq}_{\Delta_C}
\end{pmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
\Sigma^{\iota\iota}_{\Delta_C} &= nhE\begin{pmatrix}
(nh_X)^{-2}\sum_{i=1}^n \psi_2(t_\iota)\psi'_2(t_\iota), & (nh_X)^{-2}\sum_{i=1}^n \psi_2(t_\iota)\psi'_2(t_0), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_2(t_\iota)\psi'_1(t_\iota)\\
(nh_X)^{-2}\sum_{i=1}^n \psi_2(t_0)\psi'_2(t_\iota), & (nh_X)^{-2}\sum_{i=1}^n \psi_2(t_0)\psi'_2(t_0), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_2(t_0)\psi'_1(t_\iota)\\
n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_1(t_\iota)\psi'_2(t_\iota), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_1(t_\iota)\psi'_2(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_\iota)\psi'_1(t_\iota)
\end{pmatrix}\\
&= \begin{pmatrix}
C_X^{-1}\frac{1}{f(t_\iota)}\kappa E(U^2|t_\iota) & 0 & 0\\
0 & C_X^{-1}\frac{1}{f(t_0)}\kappa E(U^2|t_0) & 0\\
0 & 0 & C_t^{-1}f(t_\iota)\kappa H_{t_\iota}^{-1}E(XX'\epsilon^2|t_\iota)H_{t_\iota}^{-1}
\end{pmatrix}\\
\Sigma^{\iota\gamma}_{\Delta_C} &= nhE\begin{pmatrix}
(nh_X)^{-2}\sum_{i=1}^n \psi_2(t_\iota)\psi'_2(t_\gamma), & (nh_X)^{-2}\sum_{i=1}^n \psi_2(t_\iota)\psi'_2(t_0), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_2(t_\iota)\psi'_1(t_\gamma)\\
(nh_X)^{-2}\sum_{i=1}^n \psi_2(t_0)\psi'_2(t_\gamma), & (nh_X)^{-2}\sum_{i=1}^n \psi_2(t_0)\psi'_2(t_0), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_2(t_0)\psi'_1(t_\gamma)\\
n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_1(t_\iota)\psi'_2(t_\gamma), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_1(t_\iota)\psi'_2(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_\iota)\psi'_1(t_\gamma)
\end{pmatrix}\\
&= \begin{pmatrix}
0 & 0 & 0\\
0 & C_X^{-1}\frac{1}{f(t_0)}\kappa E(U^2|t_0) & 0\\
0 & 0 & 0
\end{pmatrix}
\end{aligned}
$$

where $\kappa = \int k^2(v)dv$,

$$E[\sum_{i=1}^n \psi_1(t_\iota)\psi'_2(t_\gamma)] = nh_X f(t_\iota)f(t_\gamma)^{-1}H_{t_\iota}^{-1}E(X\epsilon U|t_\iota)\int k(v)k(\frac{h_t v + t_\iota - t_\gamma}{h_X})dv = 0$$

$$E[\sum_{i=1}^n \psi_1(t_\iota)\psi'_1(t_\gamma)]_{\iota\neq\gamma} = nh_t f(t_\iota)H_{t_\iota}^{-1}\int X_iX'_i\epsilon_{i,t_\iota}\epsilon_{i,t_\gamma}f(X,\epsilon|t_\iota)dxd\epsilon H_{t_\gamma}^{-1}\int k(v)k(\frac{h_t v + t_\iota - t_\gamma}{h_t})dv = 0$$

$$E[\sum_{i=1}^n \psi_2(t_\iota)\psi'_2(t_\gamma)]_{\iota\neq\gamma} = nh_X f^{-1}(t_\gamma)\int U_{i,t_\iota}U_{i,t_\gamma}du\int k(v)k(\frac{h_X v + t_\iota - t_\gamma}{h_X})dv = 0$$

□

A.2 Proof of Theorem 2

The proof of Theorem 2 is quite similar to the one of Theorem 1. Before proving the main result, we first consider the following result.

Lemma 4. *Under regularity conditions stated in Lemma 1 and 2,*

$$\hat{\Delta}_S(t) - \Delta_S(t) = Bias_{\Delta_S}(t) + R_{\Delta_S}(t)W_{\Delta_S}(t) + (s.o.)$$

where

$$\begin{aligned}
Bias_{\Delta_S}(t) &= Bias_X(t_0)\left(\beta(t) - \beta(t_0)\right) + E(X'|T = t_0)\left(Bias_t - Bias_{t_0}\right) \\
R_{\Delta_S}(t) &= \left(\left(\beta(t) - \beta(t_0)\right)', \quad E(X'|T = t_0), \quad -E(X'|T = t_0)\right), \\
W_{\Delta_S}(t) &= \left(\left((nh_X)^{-1}\sum_{i=1}^{n}\psi_2(t_0)\right)', \left((nh_t)^{-1}\sum_{i=1}^{n}\psi_1(t)\right)', \left((nh_t)^{-1}\sum_{i=1}^{n}\psi_1(t_0)\right)'\right)'.
\end{aligned}$$

Proof. The proof of this result follows using essentially the same arguments as in Lemma 3. □

Proof of Theorem 2. Using the result of Lemma 4, the proof follows essentially the same arguments as in Theorem 1.

$$\sqrt{nh}\left(\hat{\Delta}_S(t_1) - \Delta_S(t_1) - Bias_{\Delta_S}(t_1), \quad \cdots, \quad \hat{\Delta}_S(t_q) - \Delta_S(t_q) - Bias_{\Delta_S}(t_q)\right)' \xrightarrow{d} N(0, \mathbb{V}_S)$$

where

$$\begin{aligned}
\mathbb{V}_S &= R_{\Delta_S}\Sigma_{\Delta_S}R'_{\Delta_S}, \\
R_{\Delta_S} &= Diag\left(R_{\Delta_S}(t_1), \quad \cdots, \quad R_{\Delta_S}(t_q)\right), \\
W_{\Delta_S} &= \left(W'_{\Delta_S}(t_1), \quad \cdots, \quad W'_{\Delta_S}(t_q)\right)', \\
\Sigma_{\Delta_S} &= E\left(nhW_{\Delta_S}W'_{\Delta_S}\right)
\end{aligned}$$

$$= nhE\begin{pmatrix} W_{\Delta_S}(t_1)W'_{\Delta_S}(t_1) & W_{\Delta_S}(t_1)W'_{\Delta_S}(t_2) & \cdots & W_{\Delta_S}(t_1)W'_{\Delta_S}(t_q) \\ W_{\Delta_S}(t_2)W'_{\Delta_S}(t_1) & W_{\Delta_S}(t_2)W'_{\Delta_S}(t_2) & \cdots & W_{\Delta_S}(t_2)W'_{\Delta_S}(t_q) \\ \vdots & \vdots & \ddots & \vdots \\ W_{\Delta_S}(t_q)W'_{\Delta_S}(t_1) & W_{\Delta_S}(t_q)W'_{\Delta_S}(t_2) & \cdots & W_{\Delta_S}(t_q)W'_{\Delta_S}(t_q) \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma^{11}_{\Delta_S} & \Sigma^{12}_{\Delta_S} & \cdots & \Sigma^{1q}_{\Delta_S} \\ \Sigma^{21}_{\Delta_S} & \Sigma^{22}_{\Delta_S} & \cdots & \Sigma^{2q}_{\Delta_S} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma^{q1}_{\Delta_S} & \Sigma^{q2}_{\Delta_S} & \cdots & \Sigma^{qq}_{\Delta_S} \end{pmatrix}$$

where

$$\Sigma^{\iota\iota}_{\Delta_S} = nhE\begin{pmatrix} (nh_X)^{-2}\sum_{i=1}^n \psi_2(t_0)\psi'_2(t_0), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_2(t_0)\psi'_1(t_\iota), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_2(t_0)\psi'_1(t_0) \\ n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_1(t_\iota)\psi'_2(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_\iota)\psi'_1(t_\iota), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_\iota)\psi'_1(t_0) \\ n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_1(t_0)\psi'_2(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_0)\psi'_1(t_\iota), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_0)\psi'_1(t_0) \end{pmatrix}$$

$$= \begin{pmatrix} C_X^{-1}\frac{1}{f(t_0)}\kappa E(U^2|t_0) & 0 & 0 \\ 0 & C_t^{-1}f(t_\iota)\kappa H_{t_\iota}^{-1}E(XX'\epsilon^2|t_\iota)H_{t_\iota}^{-1} & 0 \\ 0 & 0 & C_t^{-1}f(t_0)\kappa H_{t_0}^{-1}E(XX'\epsilon^2|t_0)H_{t_0}^{-1} \end{pmatrix}$$

$$\Sigma^{\iota\gamma}_{\Delta_S} = nhE\begin{pmatrix} (nh_X)^{-2}\sum_{i=1}^n \psi_2(t_0)\psi'_2(t_0), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_2(t_0)\psi'_1(t_\gamma), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_2(t_0)\psi'_1(t_0) \\ n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_1(t_\iota)\psi'_2(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_\iota)\psi'_1(t_\gamma), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_\iota)\psi'_1(t_0) \\ n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_1(t_0)\psi'_2(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_0)\psi'_1(t_\gamma), & (nh_t)^{-2}\sum_{i=1}^n \psi_1(t_0)\psi'_1(t_0) \end{pmatrix}$$

$$= \begin{pmatrix} C_X^{-1}\frac{1}{f(t_0)}\kappa E(U^2|t_0) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & C_t^{-1}f(t_0)\kappa H_{t_0}^{-1}E(XX'\epsilon^2|t_0)H_{t_0}^{-1} \end{pmatrix}$$

□

A.3 Proof of Theorem 3

Lemma 5. *Under regularity conditions stated in Lemma 1 and 2,*

$$\hat{\Delta}_{C_k}(t) - \Delta_{C_k}(t) = Bias_{\Delta_{C_k}}(t) + R_{\Delta_{C_k}}(t)W_{\Delta_{C_k}}(t) + (s.o.)$$

where

$$
\begin{aligned}
Bias_{\Delta_{C_k}}(t) &= \left(Bias_{X_k}(t) - Bias_{X_k}(t_0)\right)\beta(t) + \left(E(X_k|T=t) - E(X_k|T=t_0)\right) Bias_{t,k} \\
R_{\Delta_{C_k}}(t) &= \left(\beta_k(t)', \quad -\beta_k(t)', \quad \left(E(X_k|T=t) - E(X_k|T=t_0)\right)\right), \\
W_{\Delta_{C_k}}(t) &= \left(\left((nh_{X_k})^{-1}\sum_{i=1}^{n}\psi_{2,k}(t)\right)', \left((nh_{X_k})^{-1}\sum_{i=1}^{n}\psi_{2,k}(t_0)\right)', \left((nh_t)^{-1}\sum_{i=1}^{n}\psi_{1,k}(t)\right)'\right)'
\end{aligned}
$$

where k denotes the kth element.

Proof. The proof of this result follows using essentially the same arguments as in Lemma 3. □

Proof of Theorem 3. Using the result of Lemma 5, the proof follows essentially the same arguments as in Theorem 1.

$$
\sqrt{nh}\left(\hat{\Delta}_{C_k}(t_1) - \Delta_{C_k}(t_1) - Bias_{\Delta_{C_k}}(t_1), \quad \cdots, \quad \hat{\Delta}_{C_k}(t_q) - \Delta_{C_k}(t_q) - Bias_{\Delta_{C_k}}(t_q)\right)' \xrightarrow{d} N(0, \mathbb{V}_{C_k})
$$

where

$$
\begin{aligned}
\mathbb{V}_{C_k} &= R_{\Delta_{C_k}}\Sigma_{\Delta_{C_k}}R'_{\Delta_{C_k}}, \\
R_{\Delta_{C_k}} &= Diag\left(R_{\Delta_{C_k}}(t_1), \quad \cdots, \quad R_{\Delta_{C_k}}(t_q)\right), \\
W_{\Delta_{C_k}} &= \left(W'_{\Delta_{C_k}}(t_1), \quad \cdots, \quad W'_{\Delta_{C_k}}(t_q)\right)' \\
\Sigma_{\Delta_{C_k}} &= E\left(nhW_{\Delta_{C_k}}W'_{\Delta_{C_k}}\right) \\
&= nhE\begin{pmatrix}
W_{\Delta_{C_k}}(t_1)W'_{\Delta_{C_k}}(t_1) & W_{\Delta_{C_k}}(t_1)W'_{\Delta_{C_k}}(t_2) & \cdots & W_{\Delta_{C_k}}(t_1)W'_{\Delta_{C_k}}(t_q) \\
W_{\Delta_{C_k}}(t_2)W'_{\Delta_{C_k}}(t_1) & W_{\Delta_{C_k}}(t_2)W'_{\Delta_{C_k}}(t_2) & \cdots & W_{\Delta_{C_k}}(t_2)W'_{\Delta_{C_k}}(t_q) \\
\vdots & \vdots & \ddots & \vdots \\
W_{\Delta_{C_k}}(t_q)W'_{\Delta_{C_k}}(t_1) & W_{\Delta_{C_k}}(t_q)W'_{\Delta_{C_k}}(t_2) & \cdots & W_{\Delta_{C_k}}(t_q)W'_{\Delta_{C_k}}(t_q)
\end{pmatrix} \\
&= \begin{pmatrix}
\Sigma^{11}_{\Delta_{C_k}} & \Sigma^{12}_{\Delta_{C_k}} & \cdots & \Sigma^{1p}_{\Delta_{C_k}} \\
\Sigma^{21}_{\Delta_{C_k}} & \Sigma^{22}_{\Delta_{C_k}} & \cdots & \Sigma^{2p}_{\Delta_{C_k}} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma^{p1}_{\Delta_{C_k}} & \Sigma^{p2}_{\Delta_{C_k}} & \cdots & \Sigma^{pp}_{\Delta_{C_k}}
\end{pmatrix}
\end{aligned}
$$

where

$$\Sigma^{\iota\iota}_{\Delta_{C_k}} = nhE\begin{pmatrix} (nh_{X_k})^{-2}\sum_{i=1}^n \psi_{2,k}(t_\iota)\psi'_{2,k}(t_\iota), & (nh_{X_k})^{-2}\sum_{i=1}^n \psi_{2,k}(t_\iota)\psi'_{2,k}(t_0), & n^{-2}h_{X_k}^{-1}h_t^{-1}\sum_{i=1}^n \psi_{2,k}(t_\iota)\psi'_{1,k}(t_\iota) \\ (nh_{X_k})^{-2}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{2,k}(t_\iota), & (nh_{X_k})^{-2}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{2,k}(t_0), & n^{-2}h_{X_k}^{-1}h_t^{-1}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{1,k}(t_\iota) \\ n^{-2}h_{X_k}^{-1}h_t^{-1}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{2,k}(t_\iota), & n^{-2}h_{X_k}^{-1}h_t^{-1}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{2,k}(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{1,k}(t_\iota) \end{pmatrix}$$

$$= \begin{pmatrix} C_{X_k}^{-1}\frac{1}{f(t_\iota)}\kappa E(U_k^2|t_\iota) & 0 & 0 \\ 0 & C_{X_k}^{-1}\frac{1}{f(t_0)}\kappa E(U_k^2|t_0) & 0 \\ 0 & 0 & C_t^{-1}f(t_\iota)\kappa H_{t_\iota,k}^{-1}E(X_k^2\epsilon_k^2|t_\iota)H_{t_\iota,k}^{-1} \end{pmatrix}$$

$$\Sigma^{\iota\gamma}_{\Delta_{C_k}} = nhE\begin{pmatrix} (nh_{X_k})^{-2}\sum_{i=1}^n \psi_{2,k}(t_\iota)\psi'_{2,k}(t_\gamma), & (nh_{X_k})^{-2}\sum_{i=1}^n \psi_{2,k}(t_\iota)\psi'_{2,k}(t_0), & n^{-2}h_{X_k}^{-1}h_t^{-1}\sum_{i=1}^n \psi_{2,k}(t_\iota)\psi'_{1,k}(t_\gamma) \\ (nh_{X_k})^{-2}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{2,k}(t_\gamma), & (nh_{X_k})^{-2}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{2,k}(t_0), & n^{-2}h_{X_k}^{-1}h_t^{-1}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{1,k}(t_\gamma) \\ n^{-2}h_{X_k}^{-1}h_t^{-1}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{2,k}(t_\gamma), & n^{-2}h_{X_k}^{-1}h_t^{-1}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{2,k}(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{1,k}(t_\gamma) \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & C_{X_k}^{-1}\frac{1}{f(t_0)}\kappa E(U_k^2|t_0) & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where k denotes the kth element of a vector or the kth row of a matrix.

□

A.4 Proof of Theorem 4

As a first step, we prove the following lemma.

Lemma 6. *Under regularity conditions stated in Lemma 1 and 2,*

$$\hat{\Delta}_{S_k}(t) - \Delta_{S_k}(t) = Bias_{\Delta_{S_k}}(t) + R_{\Delta_{S_k}}(t)W_{\Delta_{S_k}}(t) + (s.o.)$$

where

$$\begin{aligned} Bias_{\Delta_{S_k}}(t) &= Bias_{X_k}(t_0)\left(\beta_k(t) - \beta_k(t_0)\right) + E(X_k|T=t_0)\left(Bias_{t,k} - Bias_{t_0,k}\right) \\ R_{\Delta_{S_k}}(t) &= \Big(\left(\beta_k(t) - \beta_k(t_0)\right)', \quad E(X_k|T=t_0), \quad -E(X_k|T=t_0)\Big), \\ W_{\Delta_{S_k}}(t) &= \Big(\big((nh_{X_k})^{-1}\sum_{i=1}^n \psi_{2,k}(t_0)\big)', \big((nh_t)^{-1}\sum_{i=1}^n \psi_{1,k}(t)\big)', \big((nh_{t_0})^{-1}\sum_{i=1}^n \psi_{1,k}(t_0)\big)'\Big)'. \end{aligned}$$

Proof. The proof follows essentially the same argument as in Lemma 5. □

The proof of Theorem 4. Using the result of Lemma 6, the proof follows essentially the same arguments as in Theorem 3.

$$\sqrt{nh}\left(\hat{\Delta}_{S_k}(t_1) - \Delta_{S_k}(t_1) - Bias_{\Delta_{S_k}}(t_1), \quad \cdots, \quad \hat{\Delta}_{S_k}(t_q) - \Delta_{S_k}(t_q) - Bias_{\Delta_{S_k}}(t_q)\right)' \xrightarrow{d} N(0, \mathbb{V}_{S_k})$$

where

$$
\begin{aligned}
\mathbb{V}_{S_k} &= R_{\Delta_{S_k}} \Sigma_{\Delta_{S_k}} R'_{\Delta_{S_k}}, \\
R_{\Delta_{S_k}} &= Diag\left(R_{\Delta_{S_k}}(t_1), \quad \cdots, \quad R_{\Delta_{S_k}}(t_q) \right), \\
W_{\Delta_{S_k}} &= \left(W'_{\Delta_{S_k}}(t_1), \quad \cdots, \quad W'_{\Delta_{S_k}}(t_q) \right)', \\
\Sigma_{\Delta_{S_k}} &= E\left(nhW_{\Delta_{S_k}} W'_{\Delta_{S_k}} \right) \\
&= nhE \begin{pmatrix} W_{\Delta_{S_k}}(t_1)W'_{\Delta_{S_k}}(t_1) & W_{\Delta_{S_k}}(t_1)W'_{\Delta_{S_k}}(t_2) & \cdots & W_{\Delta_{S_k}}(t_1)W'_{\Delta_{S_k}}(t_q) \\ W_{\Delta_{S_k}}(t_2)W'_{\Delta_{S_k}}(t_1) & W_{\Delta_{S_k}}(t_2)W'_{\Delta_{S_k}}(t_2) & \cdots & W_{\Delta_{S_k}}(t_2)W'_{\Delta_{S_k}}(t_q) \\ \vdots & \vdots & \ddots & \vdots \\ W_{\Delta_{S_k}}(t_q)W'_{\Delta_{S_k}}(t_1) & W_{\Delta_{S_k}}(t_q)W'_{\Delta_{S_k}}(t_2) & \cdots & W_{\Delta_{S_k}}(t_q)W'_{\Delta_{S_k}}(t_q) \end{pmatrix} \\
&= \begin{pmatrix} \Sigma^{11}_{\Delta_{S_k}} & \Sigma^{12}_{\Delta_{S_k}} & \cdots & \Sigma^{1p}_{\Delta_{S_k}} \\ \Sigma^{21}_{\Delta_{S_k}} & \Sigma^{22}_{\Delta_{S_k}} & \cdots & \Sigma^{2p}_{\Delta_{S_k}} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma^{p1}_{\Delta_{S_k}} & \Sigma^{p2}_{\Delta_{S_k}} & \cdots & \Sigma^{pp}_{\Delta_{S_k}} \end{pmatrix}
\end{aligned}
$$

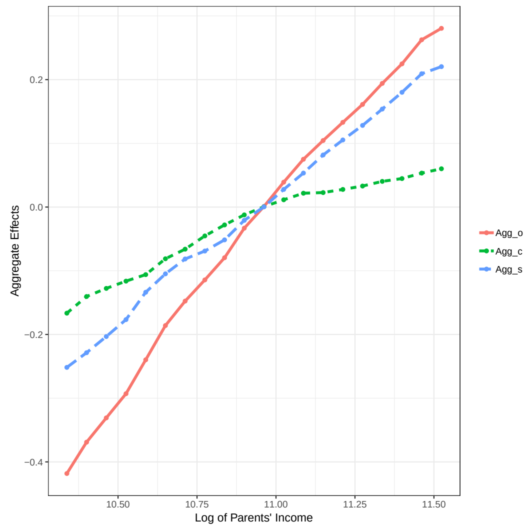
where

$$
\begin{aligned}
\Sigma^{\iota\iota}_{\Delta_{S_k}} &= nhE \begin{pmatrix} (nh_X)^{-2}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{2,k}(t_0), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{1,k}(t_\iota), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{1,k}(t_0) \\ n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{2,k}(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{1,k}(t_\iota), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{1,k}(t_0) \\ n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_{1,k}(t_0)\psi'_{2,k}(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_0)\psi'_{1,k}(t_\iota), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_0)\psi'_{1,k}(t_0) \end{pmatrix} \\
&= \begin{pmatrix} C^{-1}_{X_k}\frac{1}{f(t_0)}\kappa E(U_k^2|t_0) & 0 & 0 \\ 0 & C_t^{-1}f(t_\iota)\kappa H^{-1}_{t_\iota,k}E(X_k^2\epsilon_k^2|t_\iota)H^{-1}_{t_\iota,k} & 0 \\ 0 & 0 & C_t^{-1}f(t_0)\kappa H^{-1}_{t_0,k}E(X_k^2\epsilon_k^2|t_0)H^{-1}_{t_0,k} \end{pmatrix} \\
\Sigma^{\iota\gamma}_{\Delta_{S_k}} &= nhE \begin{pmatrix} (nh_X)^{-2}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{2,k}(t_0), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{1,k}(t_\gamma), & n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_{2,k}(t_0)\psi'_{1,k}(t_0) \\ n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{2,k}(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{1,k}(t_\gamma), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_\iota)\psi'_{1,k}(t_0) \\ n^{-2}h_X^{-1}h_t^{-1}\sum_{i=1}^n \psi_{1,k}(t_0)\psi'_{2,k}(t_0), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_0)\psi'_{1,k}(t_\gamma), & (nh_t)^{-2}\sum_{i=1}^n \psi_{1,k}(t_0)\psi'_{1,k}(t_0) \end{pmatrix} \\
&= \begin{pmatrix} C^{-1}_{X}\frac{1}{f(t_0)}\kappa E(U_k^2|t_0) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & C_t^{-1}f(t_0)\kappa H^{-1}_{t_0,k}E(X_k^2\epsilon_k^2|t_0)H^{-1}_{t_0,k} \end{pmatrix}
\end{aligned}
$$

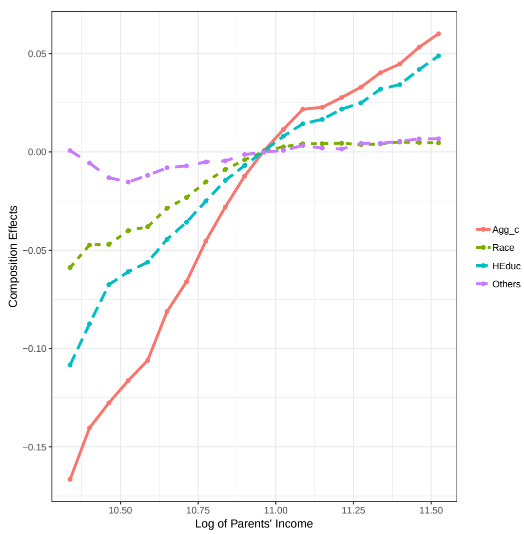
□

B Appendix 1 - Main Tables and Figures

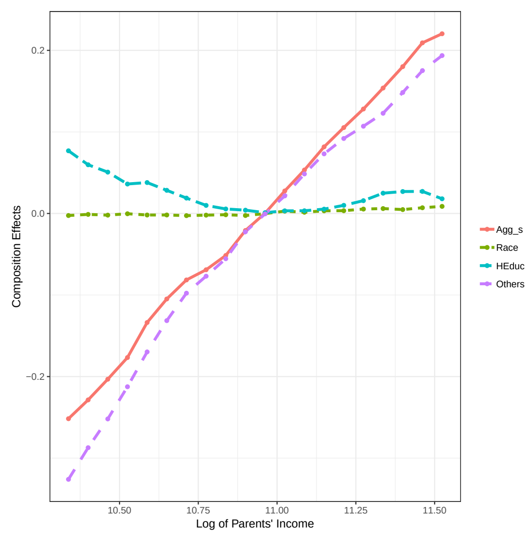
Figure 1: Decomposition across Parents' Incomes



(a) Aggregate effects



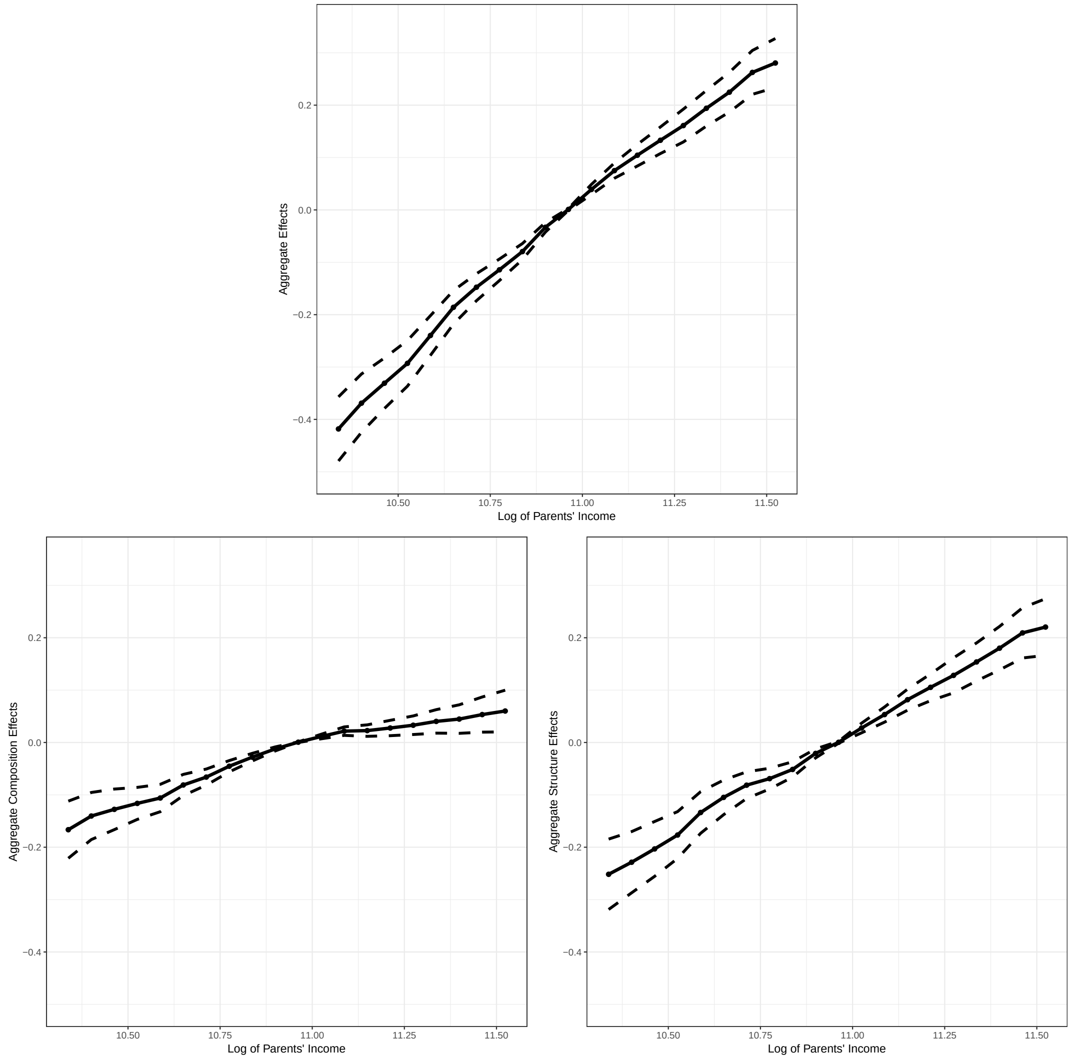
(b) Composition effects



(c) Structure effects

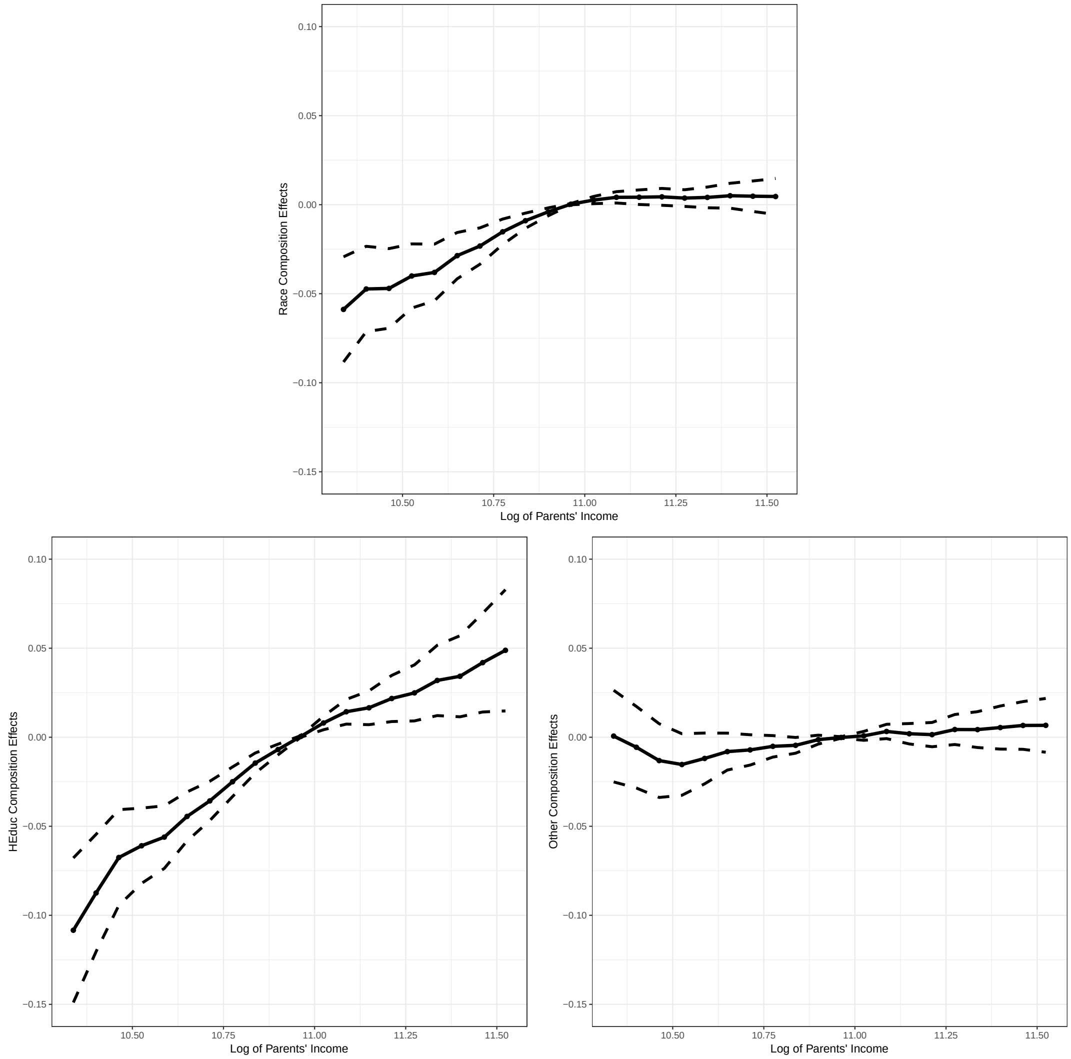
Notes: The top panel plots the aggregate aggregates across parents' incomes. The bottom left papel plots the detailed composition effects across parents' incomes. The bottom right panel plots the detailed structure effects across parents' incomes.

Figure 2: Decomposition across Parents' Incomes



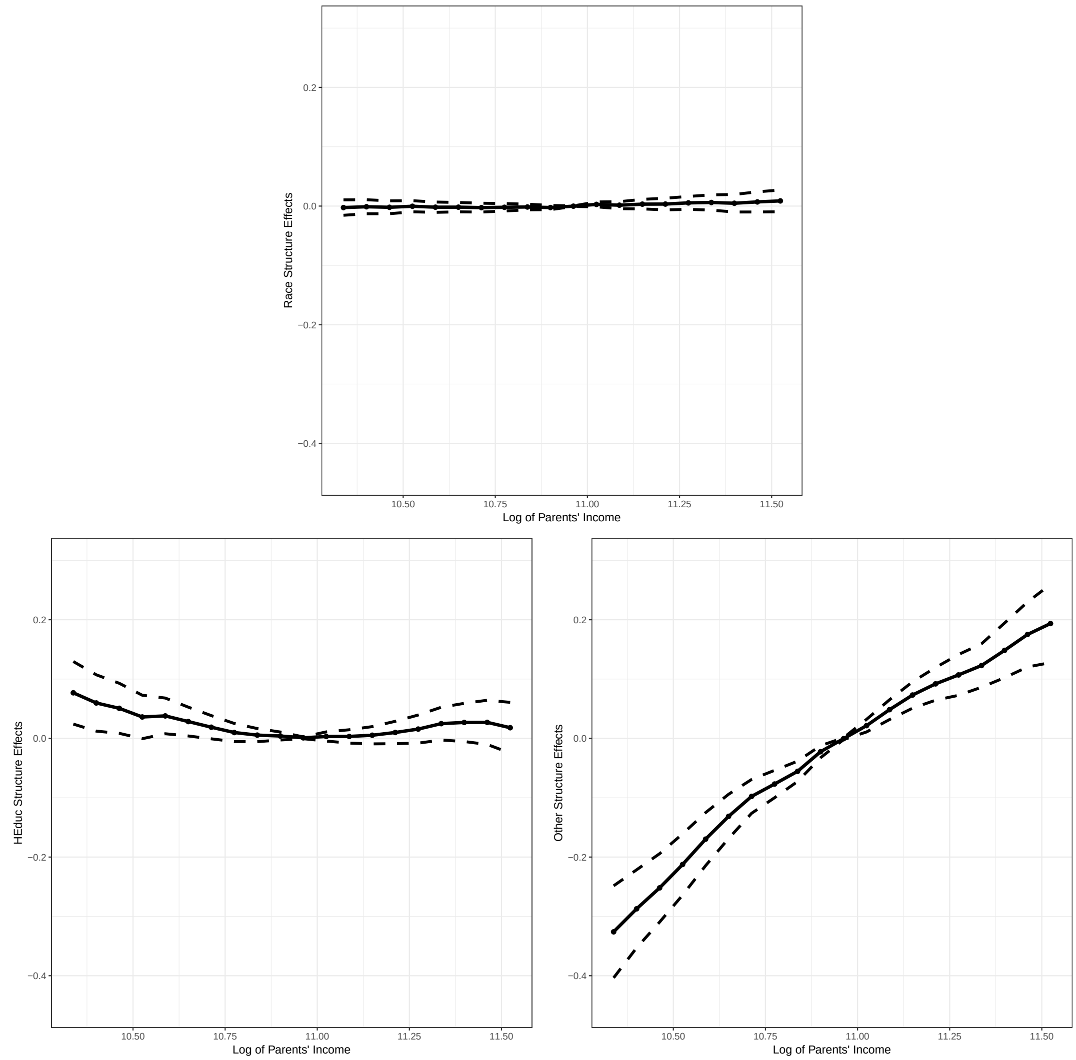
Notes: The top panel plots the overall differences across parents' incomes. The bottom left papel plots the aggregate composition effects across parents' incomes. The bottom right panel plots the aggregate structure effects across parents' incomes.

Figure 3: Decomposition across Parents' Incomes



Notes: The top panel plots the composition effects associated to race across parents' incomes. The bottom left papel plots the composition effects associated to head's education attainment across parents' incomes. The bottom right panel plots the sum of the composition effects associated to other characteristics across parents' incomes.

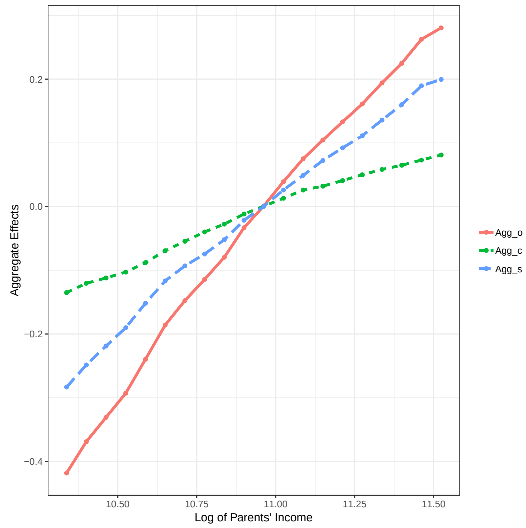
Figure 4: Decomposition across Parents' Incomes



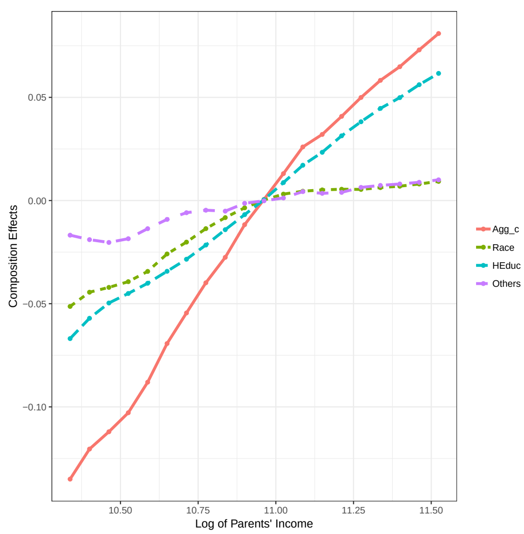
Notes: The top panel plots the structure effects associated to race across parents' incomes. The bottom left papel plots the structure effects associated to head's education attainment across parents' incomes. The bottom right panel plots the sum of the structure effects associated to other characteristics across parents' incomes.

Figure 5 to 8 plot the decomposition results as robustness check in which we switch the reference. More specifically, we use the coefficients from smooth coefficient models in which the we estimate the conditional expectation of child's income conditional on t instead of t_0.

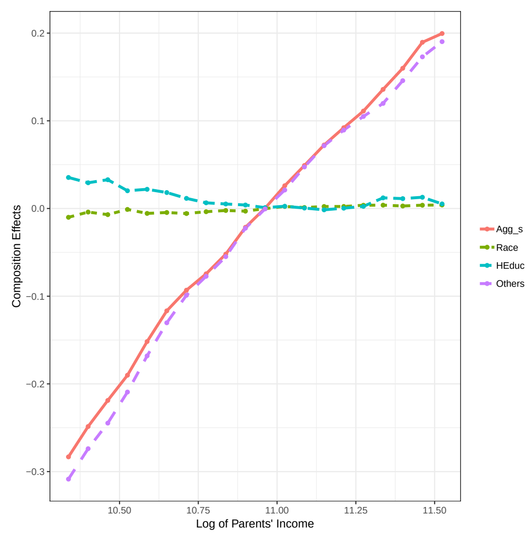
Figure 5: Decomposition across Parents' Incomes



(a) Aggregate effects



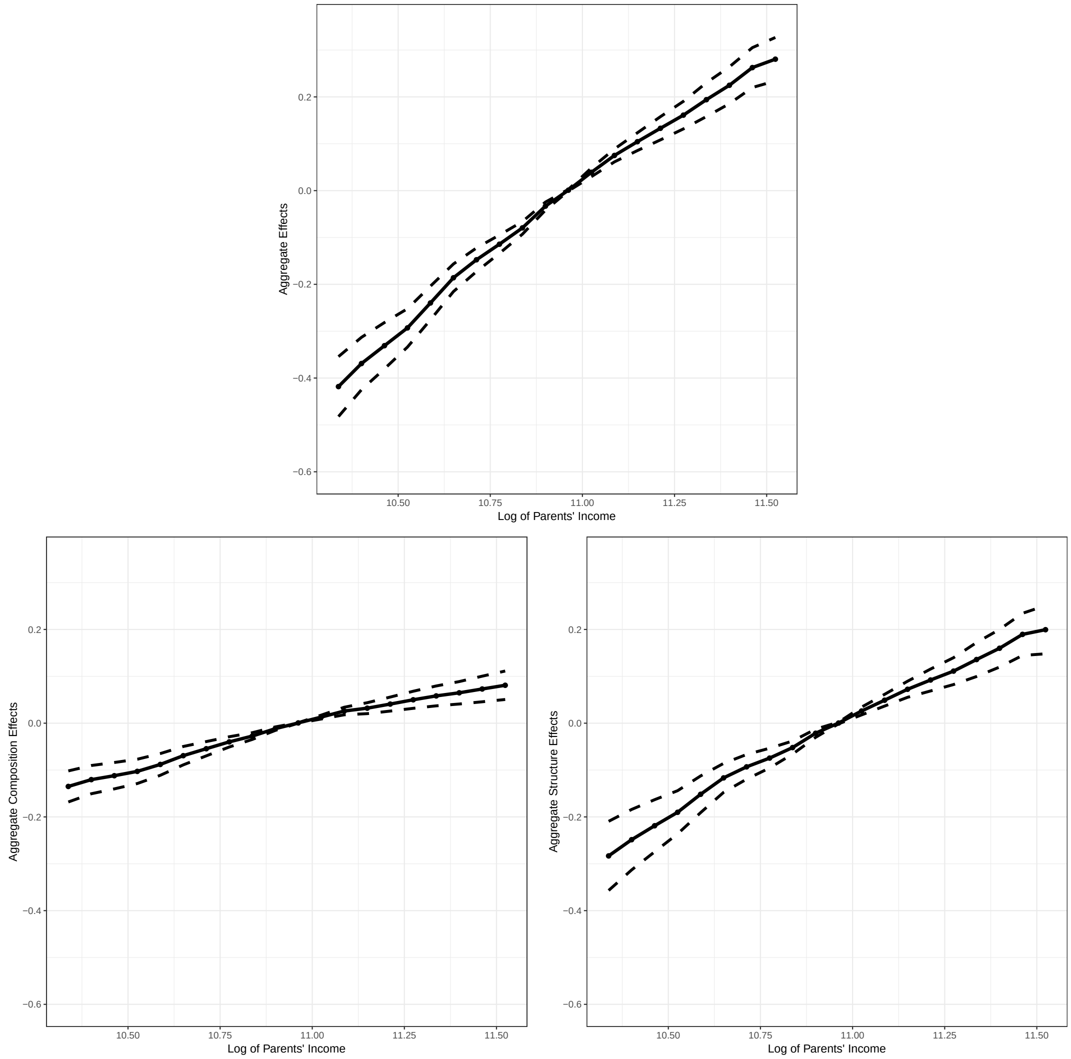
(b) Composition effects



(c) Structure effects

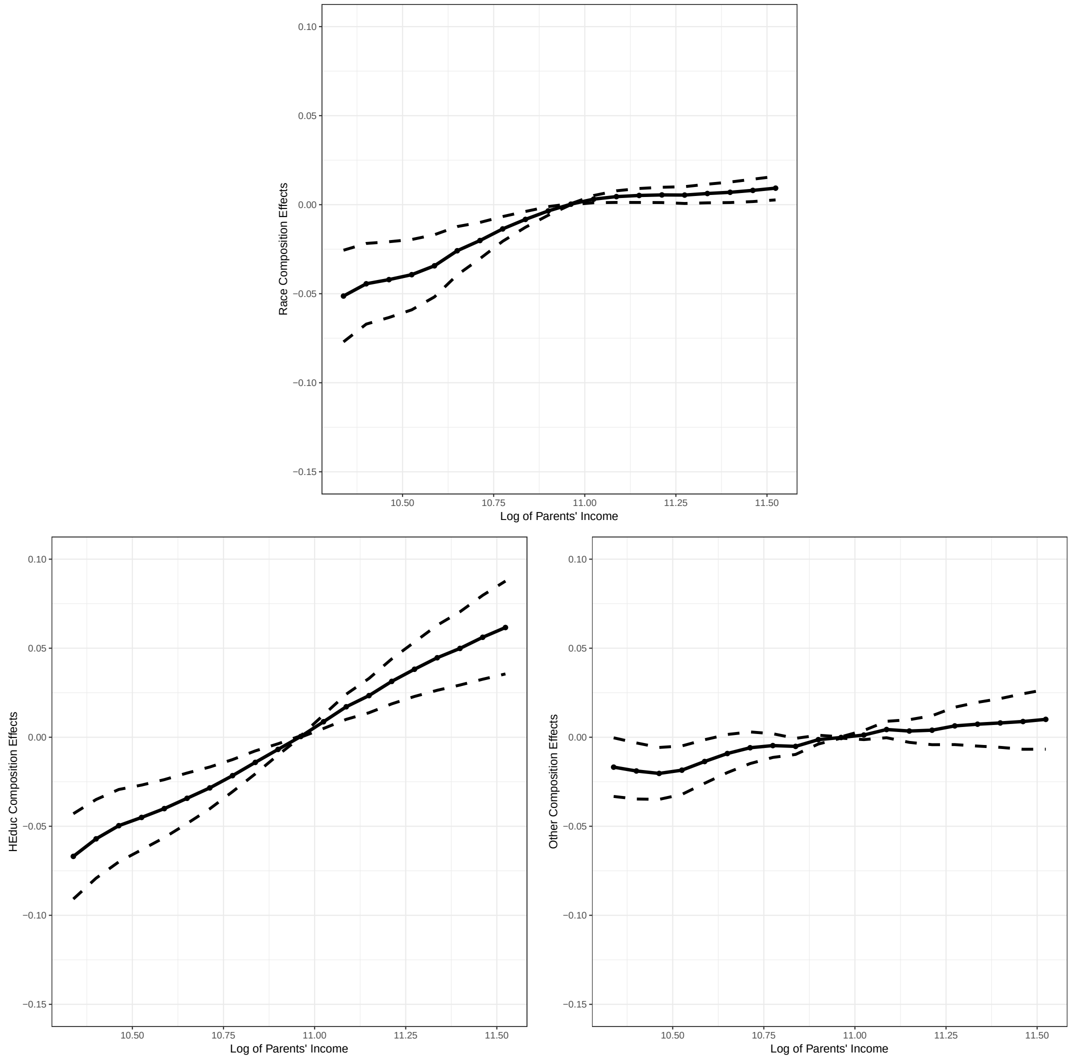
Notes: The top panel plots the aggregate aggregates across parents' incomes. The bottom left papel plots the detailed composition effects across parents' incomes. The bottom right panel plots the detailed structure effects across parents' incomes.

Figure 6: Decomposition across Parents' Incomes



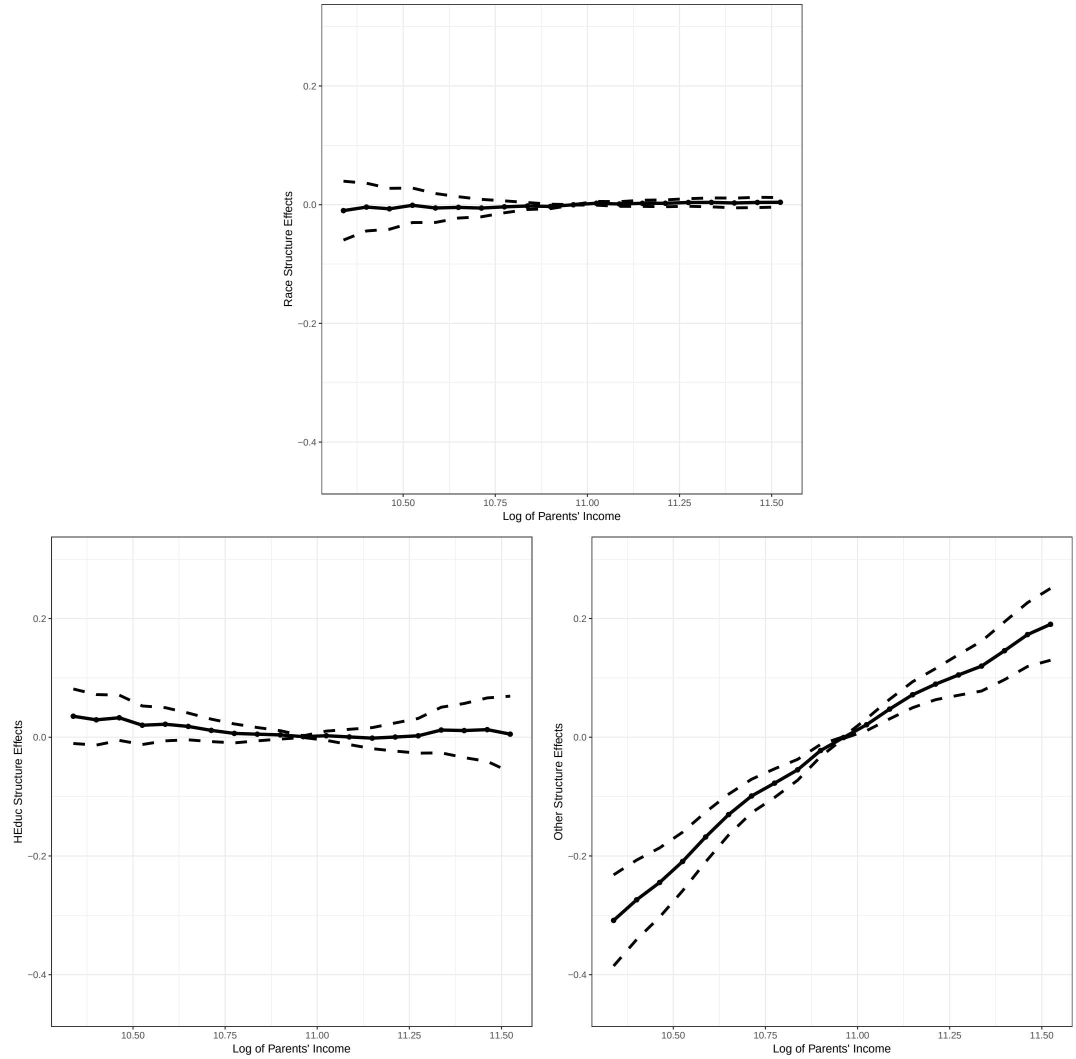
Notes: The top panel plots the overall differences across parents' incomes. The bottom left papel plots the aggregate composition effects across parents' incomes. The bottom right panel plots the aggregate structure effects across parents' incomes.

Figure 7: Decomposition across Parents' Incomes



Notes: The top panel plots the composition effects associated to race across parents' incomes. The bottom left papel plots the composition effects associated to head's education attainment across parents' incomes. The bottom right panel plots the sum of the composition effects associated to other characteristics across parents' incomes.

Figure 8: Decomposition across Parents' Incomes



Notes: The top panel plots the structure effects associated to race across parents' incomes. The bottom left papel plots the structure effects associated to head's education attainment across parents' incomes. The bottom right panel plots the sum of the structure effects associated to other characteristics across parents' incomes.