Inference on Counterfactual Transition Matrices

Brantly Callaway* Chengye Jia†

January 15, 2021

Job Market Paper

Abstract

Transition matrices provide a very useful way to summarize the dependence between two random variables; for example, the dependence between parents' income and child's income. This paper considers estimation and inference techniques for (i) conditional transition matrices – transition matrices that are conditional on some vector of covariates, (ii) counterfactual transition matrices – transition matrices that arise from holding fixed conditional transition matrices but adjusting the distribution of the covariates, and (iii) transition matrix average partial effects. Estimating conditional transition matrices is closely related to estimating conditional distribution functions, and we propose new semiparametric distribution regression estimators that may be of interest in other contexts as well. We also derive uniform inference results for transition matrices that allow researchers to account for issues such as multiple testing that naturally arise when estimating a transition matrix. We use our results to study differences in intergenerational mobility for black families and white families. In the application, we document large differences between the transition matrices of black and white families. We also show that these differences are partially, but not fully, explained by differences in the distributions of other family characteristics.

Keywords: Transition Matrix, Semiparametric Distribution Regression, Intergenerational Mobility

JEL Codes: C21, J62

*Assistant Professor, Department of Economics, University of Georgia. Email: brantly.callaway@uga.edu

†Ph.D. Student, Department of Economics, Temple University. Email: chengye.jia@temple.edu

1 Introduction

Transition matrices are useful tools to describe the dependence between two random variables. They are probably most well-known for being used to study intergenerational income mobility (Jantti et al. (2006), Bhattacharya and Mazumder (2011), Black and Devereux (2011), and Richey and Rosburg (2018), among others), but transition matrices also show up in other contexts as well such as understanding the probability of changing from the original state to the next state which is widely used as a Markov chain in macroeconomics (e.g., Bai and Wang (2011), Jones (2005), and Boreiko, Kaniovski, Kaniovski, and Pflug (2017)). In the context of intergenerational mobility, typically parents' income and child's income are divided into quartiles, and the transition matrix describes the probability that the child's income is in a particular quartile conditional on their parents' income being in a particular quartile. This provides substantially more information about the relationship between parents' income and child's income relative to common alternatives such as the Intergenerational Elasticity (IGE), which corresponds to the slope of the regression of the logarithm of child's income on the log of parents' income.

The previous discussion considers an *unconditional* transition matrix. This is the most common sort of transition matrix considered in applied work. In the current paper, we are interested in conditional transition matrices. For conditional transition matrices, each matrix cell can be represented by the probability that the child's income is in a particular quartile conditional on their parents' income being in a particular quartile *and* for some particular values of the covariates. In the context of intergenerational mobility, this sort of conditional transition matrix can be useful for studying intergenerational mobility as a function of observed characteristics such as parents' education or child's sex. More generally, conditional transition matrices can be useful in other applications as well. For example, in credit risk modeling, the conditional transition matrix can be useful as a summary of transition probabilities between different states conditional on different stress scenarios or macroeconomic conditions.

Besides our direct interest in conditional transition matrices, we are also interested in counterfactual transition matrices. These are transition matrices that arise from fixing the conditional transition matrix but varying the distribution of the covariates. For example, in the application in the paper, we consider transition matrices for black families and white families. We document that there are notable differences in the transition matrices for black and white families. We also notice that, within parents' income quartiles, there are notable differences in the distributions of other covariates, and we are interested in how much these differences can explain the differences between the transition matrices of each group. In order to do this, we construct a counterfactual transition matrix for black families where their conditional transition matrix is held constant but their distribution of covariates is changed to be the distribution of covariates for white families whose parents' income is in the same quartile.

Finally, we propose transition matrix average partial effects. These appear to be a new idea in the context of transition matrices. The idea is to study how much a particular covariate affects the probability that a child's income is in a particular quartile conditional on parents' income quartile and averaged across families.

The second main contribution is that we provide a number of new estimation results. Richey and Rosburg (2018) point out that estimating conditional transition matrices is closely related to estimating conditional distribution functions. We propose semiparametric single index estimators of conditional distribution functions. These can be seen as a generalization of distribution regression estimators (Foresi and Peracchi (1995) and Chernozhukov, Fernandez-Val, and Melly (2013)) that have been used before in the literature on estimating conditional transition matrices (Richey and Rosburg (2018)). These semiparametric distribution regression estimators are extensions of results on single index models for binary dependent variables (Klein and Spady (1993)) as well as related work including Ichimura (1993) and Ahn (1997). Relative to these papers, we face several additional challenges. The first main complication is that the cutoffs of the transition matrix (e.g., quartiles of parents' income) need to be estimated. Thus, we develop some new results on semiparametric distribution regression that account for estimating the cutoffs.[1] Second, unlike binary outcome models, transition matrices are naturally subject to multiple hypothesis testing issues – e.g., estimating a four-by-four transition matrix involves estimating twelve distribution regression (three for each quartile of parents' income and noting that the fourth one can be obtained from knowledge of the first three). This suggests that researchers ought to account for multiple testing in their inference procedure. We propose reporting transition matrices graphically and develop new results to construct uniform confidence bands (which are wider than standard pointwise confidence intervals) that account for multiple

[1]Even in the literature that proposes parametric estimators of conditional distributions to estimate the transition matrix, it does not appear that these typically account for having to estimate the cutoffs. Therefore, we also provide some results for this case as well.

hypothesis testing.

Our paper is most closely related to Bhattacharya and Mazumder (2011) and Richey and Rosburg (2018). Bhattacharya and Mazumder (2011) propose conditional transition matrices. There are two main differences between our approach and theirs. First, we propose semiparametric estimators instead of nonparametric estimators of conditional transition matrices; this is important as it is more feasible to include a relatively large number of covariates using our approach as well as including discrete covariates. The second difference is that the majority of our results integrate out some or all of the covariates rather than focusing directly on conditional transition matrices. Like Richey and Rosburg (2018), we propose estimating counterfactual transition matrices, but the particular counterfactual transition matrices that we consider are different from the ones considered in that paper: our counterfactual transition matrices arise due to differences in the distributions of covariates between two groups (in our case black and white families); Richey and Rosburg (2018) considers counterfactual transition matrices arising from replacing the distribution of characteristics in one quartile of parents' income with the distribution of characteristics from a different quartile. More importantly, we also propose semiparametric (rather than parametric) estimators of conditional transition matrices and provide a number of new results on uniform inference for transition matrices that are not previously available in the literature. Finally, we rely heavily on the observation in these papers that conditional transition matrices are closely related to conditional distribution functions.

We apply our approach to study the possible role that different distributions of covariates (these include mother's age and education, child's age and education, whether or not the child grew up in an urban location, and, in some specifications, the child's score on the Armed Services Vocational Aptitude Battery) for black and white families in the same quartile of parents' income play in explaining differences in actual transition matrices across groups. Overall, we find that differences in the distribution of covariates across races accounts for some portion of the differences between transition matrices, but a larger proportion of the difference remains unaccounted for by these covariates.

The organization of this paper is as follows. Section 2 introduces the parameters we are interested in. Section 3 describes our estimation procedure for each parameter of interest and introduces the semiparametric models that are our focus. We prove the asymptotic theory for the (counterfactual) transition matrices, coefficients, average partial effects, and uniform confidence bands in Section 4. Section 5 applies these results to study differences in transition matrices of black and white families. Section 6 concludes.

2 Parameters of Interest

Notation We use the following notation throughout the paper. Let Y^c and Y^p denote the outcomes of interest – in our application, Y^c denotes child's income and Y^p denotes parents' income. Cells in a transition matrix correspond to the probability of Y^c falling in some range conditional on Y^p falling in some range. Our results cover general cases where these ranges can essentially be arbitrary, but, for simplicity we focus on the case where there are J distinct ranges for each group. We also consider the common approach of setting the cutoffs at quantiles of the distributions of Y^c and Y^p. In particular, define a grid of equally spaced q_j with $0 = q_0 < q_1 < q_2 < \cdots < q_{J-1} < q_J = 1$. In addition, define $\tau_j^g = F_{Y^g}^{-1}(q_j)$ for $g \in \{c, p\}$. In other words, τ_j^g is cutoff value of Y^g corresponding to q_j. The most commonly encountered example of a transition matrix fits into this framework with $J = 4$ and $q_1 = 0.25$, $q_2 = 0.5$, $q_3 = 0.75$, and $q_4 = 1$ which involves creating cells in the transition matrices that are divided by the quartiles of Y^c and Y^p.

2.1 Unconditional Transition Matrices

As discussed above, the cells in a transition matrix are defined by the probability that Y^c falls into some range conditional on Y^p falling in some range. Using the notation described above, we can define the (population) value of a cell in a transition matrix by

$$
\begin{aligned}
p_{j|k} &:= P\left(q_{j-1} < F_{Y^c}(Y^c) \leq q_j \mid q_{k-1} < F_{Y^p}(Y^p) \leq q_k\right) \\
&= P\left(\tau_{j-1}^c < Y^c \leq \tau_j^c \mid \tau_{k-1}^p < Y^p \leq \tau_k^p\right) \qquad (1)
\end{aligned}
$$

Equation (1) provides two equivalent expressions for the value of the (j, k) element of a transition matrix. The first is written in terms of q_j and the ranks of Y^c and Y^p. The second is written in terms of τ_j and actual values of Y^p and Y^c. To make things concrete, consider the case where $J = 4$, $j = 2$, $k = 1$, Y^c is child's income, and Y^p is parents' income. In this case, $p_{2|1}$ is the probability that child's income is between the 25th percentile and 50th percentile in the distribution of child's income conditional on parents' income being below the 25th percentile in the distribution of parents' income. Another thing to notice is that estimating transition matrices will involve estimating distributions of Y^c and Y^p. Finally, notice that, by construction $\sum_{j=1}^{J} p_{j|k} = 1$.

2.2 Conditional Transition Matrices

Although it is much less common in applied work, it is straightforward to define conditional transition matrices as

$$\begin{aligned} p_{j|k}(x) &= P\left(q_{j-1} < F_{Y^c}(Y^c) \leq q_j \mid q_{k-1} < F_{Y^p}(Y^p) \leq q_k, X = x\right) \\ &= P\left(\tau^c_{j-1} < Y^c \leq \tau^c_j \mid \tau^p_{k-1} < Y^p \leq \tau^p_k, X = x\right) \end{aligned} \tag{2}$$

Like Equation (1), Equation (2) defines the value of a cell in a transition matrix, but this is a conditional transition matrix; i.e., the transition matrix is defined for a particular value of the covariates X. To give an example, continue to consider the case where $J = 4$, Y^c is child's income, and Y^p is parents' income. Suppose the only covariate is mother's years of education. In this case, $p_{j|k}(12)$ is the value of the (j, k) cell in a transition matrix for families where the mother has exactly 12 years of education, and $p_{j|k}(16)$ would be a cell in a transition matrix for families where the mother has exactly 16 years of education. These could differ from each other in arbitrary ways.

Another thing to notice here, is that only the outermost distribution is conditional on covariates. The cutoffs remain quartiles of the unconditional distributions of Y^c and Y^p. The reason for this is that our interest centers on probabilities of falling in cutoffs that are common for all individuals but as a function of covariates. Replacing the terms involving unconditional distributions in Equation (2) (that is, $F_{Y^c}(Y^c)$ and $F_{Y^p}(Y^p)$) with corresponding terms involving conditional distributions (e.g., $F_{Y^c|X}(Y^c|X)$ and $F_{Y^p|X}(Y^p|X)$) would introduce covariate specific cutoffs for the transition matrix. These could be quite different from each other. For example, the cutoffs for families with highly educated parents could be substantially different from the cutoffs for families with less educated parents.

Finally, notice that there is a connection between the unconditional transition matrix given in Equation (1) and the conditional transition matrix in Equation (2). In particular,

$$p_{j|k} = \int_{\mathcal{X}} p_{j|k}(x) dF_{X|\tau^p_{k-1} < Y^p \leq \tau^p_k}(x) \tag{3}$$

In other words, the unconditional transition matrix can be obtained by averaging over the conditional transition matrix in Equation (2) using the distribution of covariates among units with Y^p between τ^p_{k-1} and τ^p_k.

2.3 Transition Matrix Average Partial Effects

Next, we consider the effect of particular covariates on the transition matrix. We can define the partial effect of the ℓ-th covariate on a particular cell in the transition matrix by[2]

$$PE^{\ell}_{j|k}(x) = \frac{\partial}{\partial x_\ell} p_{j|k}(x) \tag{4}$$

$PE^{\ell}_{j|k}(x)$ is the effect of increasing x_ℓ by one unit on the probability that $\tau^c_{j-1} < Y^c \leq \tau^c_j$ conditional on $\tau^p_{j-1} < Y^p \leq \tau^p_j$ and conditional on $X = x$. To give an example in the context of intergenerational mobility, suppose that $J = 4$, X includes mother's education and child's sex. $PE^{\ell}_{1|1}(x)$ would provide the change in the probability that child's income is in the first quartile when mother's education increases by one year conditional on parents' income being in the first quartile and separately for male and female children.

The above example is simplified, and in most applications the dimension of the covariates could be substantially higher than 2. Thus, in many applications, it is attractive to average the partial effects over the distribution of the covariates; i.e., we can define the average partial effects with respect to X_ℓ as follows:

$$APE^{\ell}_{j|k} = E\left[\frac{\partial}{\partial x_\ell} p_{j|k}(X)\right] \tag{5}$$

Remark 1. Equations (4) and (5) are not unique ways to report partial effects. For example, Chernozhukov, Fernández-Val, and Luo (2018) provide a method to report the partial effects

[2]The versions of partial effects in this section are for the case where X_ℓ is continuous, but it straightforward to define analogous expressions when X_ℓ is discrete.

sorted in an increasing order and indexed by the quantiles. Some other papers focus on estimating the conditional average treatment effect (CATE) such as Abrevaya, Hsu, and Lieli (2015), Rolling, Yang, and Velez (2019), Chernozhukov and Semenova (2018), and Fan, Hsu, Lieli, and Zhang (2020). However, to our knowledge, we are the first to propose partial effects in the context of transition matrices.

2.4 Counterfactual Transition Matrices

In Equation (3), we noted the connection between conditional and unconditional transition matrices. Following a large literature in econometrics on counterfactual distributions, we define a counterfactual transition matrix by considering alternative distributions of X while holding fixed the conditional transition matrix (i.e., by replacing the distribution of covariates in Equation (3) with another distribution). That is, we define

$$\theta_{j|k} = \int p_{j|k}(x)\, d\tilde{F}_X(x)$$

where $\tilde{F}_X$ is some alternative distribution for the covariates, and $\theta_{j|k}$ provides the value of a cell in a *counterfactual* transition matrix. In the context of intergenerational income mobility, there are a number of potentially interesting counterfactual transition matrices. We give some examples next.

Example 1. Consider the case where a researcher is interested in differences in intergenerational mobility for black and white families. Let $p^d_{j|k}$ denote cells in the unconditional transition matrix and $p^d_{j|k}(x)$ denote cells in a conditional matrix for families in group d for $d \in \{\text{black families}, \text{white families}\}$. Then, one potentially interesting counterfactual transition matrix is given by

$$\theta^d_{j|k} = \int p^d_{j|k}(x)\, dF_{X|\tau^p_{k-1} < Y^p < \tau^p_k, D}(x|d')$$

where d' denotes the other group. In other words, cells in this counterfactual transition matrix arise from taking the conditional transition matrix for group d but averaging over the distribution of covariates of group d' among units whose parents' income is in the same quartile. Note that these arguments also apply more generally to any case where units can be split into two groups;

for example, one could construct counterfactual transition matrices for different cohorts which is related to decompositions in Richey and Rosburg (2017) and Davis and Mazumder (2018).

Example 2. Another interesting counterfactual distribution is one where the distribution of covariates is equalized across quartiles of parents' income. This sort of counterfactual transition matrix is given by

$$\theta_{j|k} = \int p_{j|k}(x) \, dF_X(x)$$

A related counterfactual transition matrix is one where the distribution of covariates for units with $\tau^p_{k-1} < Y^p \leq \tau^p_k$ is replaced with the distribution of covariates for units with $\tau^p_{k'-1} < Y^p \leq \tau^p_{k'}$ for $k' \neq k$ (e.g., in the context of intergenerational mobility, this would correspond to replacing the distribution of covariates for families with parents' income in the first quartile with the distribution of covariates for families with parents' income in the fourth quartile). This is the sort of counterfactual distribution considered in Richey and Rosburg (2018). In this case, the counterfactual transition matrix is given by

$$\theta^d_{j|k} = \int p^d_{j|k}(x) \, dF_{X|\tau^p_{k'-1} < Y^p < \tau^p_{k'}}(x)$$

In the application in the current paper, we focus on counterfactual distributions as in Example 1.

3 Estimation

3.1 Estimating Conditional Transition Matrices

As pointed out in Richey and Rosburg (2018), there is an interesting connection between conditional transition matrices and conditional distributions. In particular, notice that we can rewrite Equation (2) as

$$p_{j|k}(x) = F_{Y^c|X,Y^p}(\tau^c_j|x,k) - F_{Y^c|X,Y^p}(\tau^c_{j-1}|x,k) \tag{6}$$

where $F_{Y^c|X,Y^p}(\tau_j^c|x,k) := P(Y^c \leq \tau_j^c|\tau_{k-1}^p < Y^p \leq \tau_k^p, X = x)$. Equation (6) suggests that estimating conditional transition matrices comes down to estimation of conditional distributions. Notice that $F_{Y^c|X,Y^p}$ is directly identified from the sampling process, so, in principle, one could estimate each conditional distribution above nonparametrically (this is the approach in Bhattacharya and Mazumder (2011)). However, this approach would suffer from the curse of dimensionality, and, in many cases, is also simply infeasible. Instead, we make the following assumption.

Assumption 1 (Single Index Conditional Distribution)**.**

$$F_{Y^c|X,Y^p}(\tau_j^c|x,k) = G_{j|k}(x'\beta_{j|k})$$

We further consider two cases of Assumption 1:

Assumption 2. *One of the following two cases holds:*

1. $G_{j|k}$ *is known*
2. $G_{j|k}$ *is unknown*

Under Assumption 2[1], the problem of estimating conditional distributions reduces to a distribution regression problem (see Chernozhukov, Fernandez-Val, and Melly (2013) for recent work on distribution regression generally and Richey and Rosburg (2018) and Callaway and Huang (2020) for applying distribution regression in the context of intergenerational mobility).

The second case corresponds to distribution regression with an unspecified link function. Klein and Spady (1993) presents a semiparametric maximum likelihood estimation for binary dependent variables. Blundell and Powell (2004) propose semiparametric methods for estimating binary response (binary choice) models with continuous endogenous regressors. Rothe (2009) proposes a semiparametric maximum likelihood method to conduct the estimation of the coefficients of a single index binary choice model with endogenous regressors. Unlike ours, this paper focuses on the coefficients of the single-index model. Rothe (2010) uses nonparametric kernel method to estimate the conditional distribution function. The model in our paper is closely connected to Klein and Spady (1993). But unlike their paper, the dependent variables are not binary and thus have to be transformed into binary variables.

3.2 Semiparametric single-index binary response model

In this section, we propose a semiparametric model to estimate the (counterfactual) conditional transition matrix. The advantages of our approach are that (i) this method does not require one to choose the link function which could be difficult in practice and, and it is not clear how one could make this choice in most applications and (ii) it does not suffer from the curse of dimensionality which would be the case for nonparametric estimators.

To start with, in order to reduce notation, we first transform the child's and parents' income into binary variables which are equal to 1 if individual income is less than the defined threshold and zero otherwise. The thresholds are equal to the unique values τ_j^g defined above. Then, we can use the semiparametric single-index binary response model to calculate the probability of child's income falling a particular range conditional on parents' income falling in a particular range. We introduce the following notation. Let

$$y_j^c = \begin{cases} 1 & Y^c \in (\tau_{j-1}^c, \tau_j^c] \\ 0 & \text{Otherwise} \end{cases} \quad \text{and} \quad y_k^p = \begin{cases} 1 & Y^p \in (\tau_{k-1}^p, \tau_k^p] \\ 0 & \text{Otherwise} \end{cases}$$

Then given the parents' income, we can write

$$P(y_j^c = 1 | x, y_k^p = 1) = G_{j|k}(x'\beta_{j|k}) - G_{j-1|k}(x'\beta_{j-1|k}) \tag{7}$$

where y_j^c and y_k^p are equal to 1 if $y_i^c \in (\tau_{j-1}^c, \tau_j^c]$ and $y_i^p \in (\tau_{k-1}^p, \tau_k^p]$ and $\beta_{j|k}$ are the coefficients when child's income and parents' income are in the jth and kth quartiles of their income distributions respectively where the equality in Equation (7) holds by Assumption 1. $G_{j|k}(\cdot)$ is the link function defined in Assumption 2 which maps the single index, $x'\beta_{j|k}$, into the probability of $y_j^c = 1$ conditional on covariates x and parents' income is in the kth quartile of their income distribution.

In this paper, we focus on comparing transition matrices for black and white families. Therefore, we somewhat specialize our notation to this case with the primary goal of making it easier to follow. Let D define whether a family is black or white; i.e., we set $D = b$ for black families and $D = w$ for white families. Then, we can define a counterfactual transition matrix for black families if they had the same distribution of covariates as white families with parents' income in the same quartile with cells that are given by

$$\theta^b_{j|k} = \int_{\mathcal{X}} P(y^c_j = 1 \mid y^p_k = 1, X = x, D = b)\, dF_{X|D=w,y^p_k=1}(x) \tag{8}$$

$\theta^b_{j|k}$ is a main parameter of interest in the application. We also compare $\theta^b_{j|k}$ to $p^b_{j|k}$ (the actual transition matrix for black families). This sort of comparison is similar to a composition effect in the decomposition literature – i.e., it is due to differences in the distribution of covariates between groups. Further, we can compare $\theta^b_{j|k}$ to $p^w_{j|k}$ (the actual transition matrix of white families). This comparison is similar to a structure effect in the decomposition literature – i.e., it is due to unexplained (by covariates) differences in transition matrices across groups.

3.3 Estimation Procedure

Now the problem is how we estimate Equations (7) and (8). In this part, we build on the semiparametric maximum likelihood method proposed in Klein and Spady (1993) to estimate the coefficient $\beta^d_{j|k}$ and the unknown link function $G^d_{j|k}$ for $d \in \{b, w\}$. Since we concentrate on estimating counterfactual distributions for black families, we focus our estimation results on estimating $\beta^d_{j|k}$ and $G^d_{j|k}$, but note that our estimation results would follow analogously for white families or, in cases where there were not two groups, for the entire population.

If the coefficients were known, we can estimate the conditional probability for the group b using the leave-one-out kernel method, i.e.,

$$\hat{G}^b_{j|k}(x'\beta^b_{j|k}) = (N^b_k h^b_{j|k})^{-1} \sum_{i \neq m}^{N^b_k} \frac{y^c_j K\left(\frac{x'_i\beta^b_{j|k} - x'_m\beta^b_{j|k}}{h^b_{j|k}}\right)}{(N^b_k h^b_{j|k})^{-1} \sum_{i \neq m}^{N^b_k} K\left(\frac{x'_i\beta^b_{j|k} - x'_m\beta^b_{j|k}}{h^b_{j|k}}\right)} \tag{9}$$

where N^b_k and $h^b_{j|k}$ represent the total number of black families and bandwidth for families with child's income in jth quantile and parents' income is in kth quantile, and where i indexes a particular family. In order to avoid having too small values of the denominator of eq. (9), we trim out small values of the denominator and restrict covariates X to a fixed set. Set the denominator as $p_{-m}(x'_m\beta) = (N^b_k h^b_{j|k})^{-1} \sum_{i \neq m}^{N^b_k} K\left(\frac{x'_i\beta^b_{j|k} - x'_m\beta^b_{j|k}}{h^b_{j|k}}\right)$ where $p_{-m}(\cdot)$ means that we leave the individual m out of sample when we calculate probability of the individual m. Denote

A_x and A_{nx} as follows:

$$A_x = \left\{x : p_{-m}(x_m'\beta) \geq \delta, \text{for all } \beta \text{ in } \mathcal{B}\right\}$$

and

$$A_{nx} = \{x : \|x - x^*\| \leq 2h, \text{for some } x^* \in A_x\}$$

where δ>0 is a constant, $\mathcal{B}$ is a compact subset in $\mathbb{R}^k$ and h is the bandwidth. A_{nx} includes A_x, which will shrink to A_x as $n \to \infty$ and $h \to 0$. With the restriction of the support of the covariates, we need to rewrite equation eq. (9) as

$$\hat{G}^b_{j|k}(x'\beta^b_{j|k}) = (N^b_k h^b_{j|k})^{-1} \sum_{i \neq m}^{N^b_k} \frac{\mathbb{I}_{nx_i} y^c_j K\left(\frac{x_i'\beta^b_{j|k} - x_m'\beta^b_{j|k}}{h^b_{j|k}}\right)}{(N^b_k h^b_{j|k})^{-1} \sum_{i \neq m}^{N^b_k} \mathbb{I}_{nx_i} K\left(\frac{x_i'\beta^b_{j|k} - x_m'\beta^b_{j|k}}{h^b_{j|k}}\right)} \tag{10}$$

where $\mathbb{I}_{nx_i} = \mathbf{1}(x_i \in A_{nx})$ is the indicator function and equal to one when the condition in the parentheses is satisfied and zero otherwise.

Next, we will use maximum likelihood method to estimate the coefficients $\beta^b_{j|k}$. The log-likelihood function is given by

$$\mathcal{L}_{j|k} = \sum_i^{N^b_k} \mathbb{I}_{x_i} \left[y^c_j \ln(\hat{G}^b_{j|k}(x'\beta^b_{j|k})) + (1 - y^c_j) \ln(1 - \hat{G}^b_{j|k}(x'\beta^b_{j|k}))\right] \tag{11}$$

where $\mathbb{I}_{x_{c'}} = \mathbf{1}(x_{c'} \in A_x)$. Then the estimate of $\beta^b_{j|k}$ can be obtained by maximizing the log-likelihood function. As in Klein and Spady (1993), the convergence rate of $G(\cdot)$ is slower than that of the coefficients though we show that the $\hat{\beta}$ is $\sqrt{n}$-consistent and has an asymptotic normal distribution in the next section.

After obtaining the estimates of coefficients and link function, we can estimate the transition matrix by using following equation:

$$\hat{p}^b_{j|k} = \frac{1}{N^b_k} \sum_i \hat{G}_{j|k}(x_i \hat{\beta}^b_{j|k})$$

The estimated counterfactual transition matrix can be obtained by using empirical distribution function such as:

$$\hat{\theta}^b_{j|k} = \frac{1}{N^w_k}\sum_{m=1}^{N^w_k} \frac{\sum_i y^c_{jb} K\left(\frac{x'_{ib}\hat{\beta}^b_{j|k} - x'_{mw}\hat{\beta}^b_{j|k}}{\hat{h}^b_{j|k}}\right)}{\sum_i K\left(\frac{x'_{ib}\hat{\beta}^b_{j|k} - x'_{mw}\hat{\beta}^b_{j|k}}{\hat{h}^b_{j|k}}\right)}$$

where N^w_k is the total number of white families with parents' income in the kth quartile, and x_{mw} is the characteristics of a white family m. $\hat{\beta}^b_{j|k}$ is estimated by using the sample of black families and Equation (11). $\hat{\theta}^b_{j|k}$ is an estimate of the (j,k) cell in the counterfactual transition matrix for black families.

Finally, the average partial effects of the ℓth covariate variable (for black families) can be estimated by using the following equation:

$$\begin{aligned}\widehat{APE}^d_{j|k} =& \frac{1}{N^b_k}\sum_m^{N^b_k} \frac{\left[\sum_i^{N^d_k} y^c_j \partial_{x_{dm}} K\left(\frac{x'_{ib}\hat{\beta}^b_{j|k} - x'_{mb}\hat{\beta}^b_{j|k}}{\hat{h}^b_{j|k}}\right)\right] \sum_j^{N^d_k} K\left(\frac{x'_{ib}\hat{\beta}^b_{j|k} - x'_{mb}\hat{\beta}^b_{j|k}}{\hat{h}^b_{j|k}}\right)}{\left[\sum_j^{N^d_k} K\left(\frac{x'_{ib}\hat{\beta}^b_{j|k} - x'_{mb}\hat{\beta}^b_{j|k}}{\hat{h}^b_{j|k}}\right)\right]^2} \\ &- \frac{1}{N^b_k}\sum_m^{N^b_k} \frac{\left[\sum_j^{N^d_k} y_j K(\frac{x'_{ib}\hat{\beta}^b_{j|k} - x'_{mb}\hat{\beta}^b_{j|k}}{\hat{h}^b_{j|k}})\right] \sum_j^{N^d_k} \partial_{x_{di}} K\left(\frac{x'_{ib}\hat{\beta}^b_{j|k} - x'_{mb}\hat{\beta}^b_{j|k}}{\hat{h}^b_{j|k}}\right)}{\left[\sum_j^{N^d_k} K\left(\frac{x'_{ib}\hat{\beta}^b_{j|k} - x'_{mb}\hat{\beta}^b_{j|k}}{\hat{h}^b_{j|k}}\right)\right]^2} \end{aligned} \tag{12}$$

where $\partial_{x_{dm}}$ is the partial derivative of kernel function with respect to the d-th variable of individual m.

4 Asymptotic properties

In this part, we derive the asymptotic theory of our estimators. We first state the assumptions and then obtain the results on consistency and asymptotic normality. Finally, we conduct inference by building uniform confidence bands and discuss the validity of the bootstrap.

4.1 Assumptions

First, we introduce some notation. We can write the counterfactual transition matrix as a map from a conditional transition matrix and a distribution of covariates as follows

$$\theta_{j|k}(p,F) = \phi(p_{j|k}(x), F_X) := \int p_{j|k}(x)\, dF_X(x)$$

Using this notation, the particular counterfactual transition matrix that we consider is

$$\theta^b_{j|k}(p,F) = \phi(p^b_{j|k}(x), F_{X|k,w})$$

where $F_{X|k,w}(x) := F_{X|\tau^p_{k-1}<Y^p\leq\tau^p_{k-1},D=w}(x)$.

In addition, in applications, all of the τ^c_j and τ^p_k need to be estimated, and we denote their estimators by $\hat{\tau}^c_j$ and $\hat{\tau}^p_k$.

To derive the limiting distribution of the counterfactual transition matrix, we make the following assumptions.

Assumption 3. *For all j and k, the space $\mathcal{B}$ of vector $\beta_{j|k}$ is compact, the true value $\beta^0_{j|k}$ is in the interior of $\mathcal{B}$, the space $\mathcal{T}$ of vector τ is compact where the true value τ^0 is in the interior of $\mathcal{T}$, single index $z_{j|k} = x^T\beta_{j|k} \in \mathcal{Z}$ and $\mathcal{Z}$ is compact and the true value $z^0_{j|k}$ is in its interior.*

Assumption 4. *Here we will omit the subscript j and k for simplicity: e.g $\tau_1 = \tau^c = \tau^c_j$ and $\tau_0 = \tau^p = \tau^p_k$. Let $D = b$ for the black families and $D = w$ for the White families. We will omit D when $D = b$. The conditional distribution $g^D(\tau, z)\colon \mathbb{R}\times\mathbb{R} \mapsto [0,1]$ is the distribution of τ^c conditional on τ^p and z, and depends on the single index z and $\tau = (\tau^c, \tau^p)$. $f^D(z)$ is the probability density of z. $f^D(z,\tau^p)$ is the joint density function of z and τ^p. $\phi^D(\tau^c,\tau^p,z) = Pr\,(Y^c_i \leq \tau^c, Y^p_i \leq \tau^p|\, z) = g^D_{\tau z}/f^D(z)$ where $g^D_{\tau z} = g^D(\tau,z)f^D(z,\tau^p)$. g^D, $g^D_{\tau z}$, $f^D(z)$ and $f^D(z,\tau_0)$ are all r times differentiable with respect to X, Z, β, τ and their derivatives are uniformly bounded on x, z and τ respectively.*

Assumption 5. *(i) The data $\{(Y^c_i, Y^p_i, X_i) : i = 1, 2, \cdots, N\}$ are i.i.d. X contains at least one continuous variable and there is no multicollinearity between covariates. (ii) The support of Black family's covariates X_b should include or at least intersect with the support of White family's covariates X_w: $X_w \subseteq X_b$ or $X_b \cap X_w \neq \emptyset$. Also for any $x \in X_b$ or X_w, the probability functions $f^D(z)$ and $f^D(z,\tau_0)$ in Assumption 4 are bouned away from zero.*

Assumption 6. *The kernel functions:* $K\colon \mathbb{R}^d \mapsto \mathbb{R}$ *satisfies the following properties: (1)* $\int K(z)dz = 1$*. (2)* $\int z^{\mu} K(z)dz = 0$ *for all* $|\mu| = 1, 2, \cdots, r-1$*. (3)*$\int |z^{\mu} K(z)|dz < \infty$ *for* $|\mu| = r$ *(4)* $K^r(z)$ *are the r-times derivatives and uniformly continuous and bounded.*

Assumption 7. *The bandwidth* $h = h_n$ *satisfies* $h \to 0$*,* $\sqrt{n}h/log(n) \to \infty$ *and* $n^{1/2}h^r \to 0$*.* $h \sim n^{-\delta}$*, where* $\delta \in (1/2r, 1/6d)$*,where d is the dimension of z.*

Assumption 3 are standard conditions in the semiparametric literature. Assumption 5 define a trimming set A_x in order to keep the denominator close to zero as n goes to infinity. Assumption 4 puts smoothness restrictions on the functions estimated by nonparametric methods. Assumption 6 defines the uniform boundedness and continuity of kernel which is of order r and r times differentiable and is a standard condition on the kernel function in the nonparametric literature. Assumption 7 sets the rate at which the bandwidth will converge to zero as n goes to infinity; this choice of the bandwidth leads to undersmoothing and thus no bias term will show up in the limiting distribution. In this paper, $3d < r$ and $d = 1$.

4.2 Asymptotic Normality

We first introduce some notations. Let $Z_{j|k} = X^T \beta_{j|k} = X^T \beta_{j|k}$ for group b and $Z^w = Z^w_{j|k} = X_w^T \beta_{j|k}$ where X_w is the vector of covariates from group w. f^w is the probability density of group w. Next, by Lemma 6,

$$\sqrt{n}\left(\hat{\tau}_j^c - \tau_j^c\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{ijc} + o_p(1) \quad \text{and} \quad \sqrt{n}\left(\hat{\tau}_k^p - \tau_k^p\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{ikp} + o_p(1)$$

where

$$\psi_{ijc} = \frac{1\{Y_i^c \leq \tau_j^c\} - F_{Y^c}(\tau_j^c)}{f_{Y^c}(\tau_j^c)} \quad \text{and} \quad \psi_{ikp} = \frac{1\{Y_i^p \leq \tau_k^p\} - F_{Y^p}(\tau_k^p)}{f_{Y^p}(\tau_k^p)}$$

Theorem 1. *Under Assumptions 1 to 7,*

$$\sqrt{n}(\hat{\theta}^b - \theta^b) \rightsquigarrow \mathbb{Z}$$

where $\mathbb{Z}$ *is a mean zero Gaussian process indexed by J with covariance function:* $E(V_{\theta,\tau_J}V_{\theta,\tau_{J'}})$

$$V_{\theta,\tau_J} = \Psi_1 + \Psi_2 + \Psi_3$$

where $V(Z, Z^w, \tau_k^p) = \frac{1\{Y^p \leq \tau_k^p\} f^w(z,\tau_k^p)}{f(z,\tau_k^p)^2 f^w(\tau_k^p)}$

$\Psi_1 = E\left(\partial_\beta G(\tau_{jk}, Z) + G(\tau_{jk}, Z)\frac{\partial_\beta f(z,\tau_k^p)}{f(z,\tau_k^p)}\right)(\hat{\beta} - \beta)$

$\Psi_2 = E(\partial_{\tau^c} G(\tau_{jk}, Z))\psi_{ijc} + \left(E(\partial_{\tau^p} G(\tau_{jk}, Z)) + E\left(G(\tau_{jk}, Z)\frac{\partial_{\tau^p} f(Z,\tau_k^p)}{f(Z,\tau_k^p)}\right)\right)\psi_{ikp}$

$\Psi_3 = \frac{\mathbb{1}\{d=w\}}{P(D=w|k)}\left(G(\tau_{jk}, Z^w) - E[G(\tau_{jk}, Z)] + (1\{Y^c \leq \tau_j^c\} - G(\tau_{jk}, Z))V(Z, Z^w, \tau_k^p)\right)$

where the $\tau_J = (\tau_j^c, \tau_k^p)$ and $J = (j, k)$ for $j = 1, \cdots, J$ and $k = 1, \cdots, J$. The proof of Theorem 1 is provided in Appendix B. Theorem 1 provides an asymptotically representation for our estimator of a particular cell in the counterfactual transition matrix. This is the key building block for developing uniform confidence bands for the transition matrix that account for multiple hypothesis testing. Next, we provide the limiting distribution for our estimators of $\beta_{j|k}$.

Theorem 2. *Assumptions 1 to 7,*

$$\sqrt{n}(\hat{\beta} - \beta) \rightsquigarrow \mathbb{B}$$

$\mathbb{B}$ *is the mean zero Gaussian distribution indexed by J with covariance function:* $E(V_{\beta,J}, V'_{\beta,J'})$ *where* $V_{\beta,J} = \Sigma^{-1}(\mathbf{M}_1 + \mathbf{M}_2)$ *and*

$$\mathbf{M}_1 = \frac{[\partial_\beta G(\tau_J, Z) - E(\partial_\beta G(\tau_{jk}, Z)|Z)]\,(1\{Y^c \leq \tau_j, Y^p \leq \tau_k\} - G(\tau_J, Z))}{f(\tau_{jk})^2 G(\tau_J, Z)\,(1 - G(\tau_J, Z))}$$

$$\mathbf{M}_2 = \Psi_p\psi_p + \Psi_c\psi_c$$

$$\Sigma = E\left(\frac{[\partial_\beta G(\tau_J, Z)]\,[\partial_{\beta'} G(\tau_J, Z)]}{G(\tau_J, Z)\,(1 - G(\tau_J, Z))}\right)$$

$$\Psi_m = E\left[\frac{(\partial_\beta G(\tau_J, Z) - E(\partial_\beta G(\tau_J, Z)|z))\,\partial_{\tau_0} G(\tau_J, Z)}{G(\tau_J, Z)(1 - G(\tau_J, Z))}\right]\psi_m \quad, m = c \text{ or } p$$

The proof of Theorem 2 is provided in Appendix B. Next, we provide the limiting distribution of our estimators of the average partial effects of the vector of covariates X.

Proposition 1. *Under Assumptions 1 to 7,*

$$\sqrt{n}(\widehat{APE} - APE) \xrightarrow{d} N(0, V_{APE})$$

with covariance function $E(V_{APE_J} V'_{APE_{J'}})$ *and* $V_{APE_J} = E(\partial_X G(\tau, Z))\hat{\beta} - \beta$

The proof of Proposition 1 is provided in Appendix B.

4.3 Inference

From Theorems 1 and 2, we can conduct point-wise inference on the parameters of interest such as each element of (counterfactual) transition matrix, the coefficient and the average partial effects, by standard Gaussian approximation. A first disadvantage of proceeding this direction is that the expressions for the asymptotic variance are complicated to estimate. In particular, note that estimating the asymptotic variance depends on estimating densities which will introduce a number of additional challenges such as choosing additional bandwidths. Instead of taking this route, we propose to conduct inference using the bootstrap. Below, we use $*$ to represent the bootstrap estimator, and the next result shows the validity of using the empirical bootstrap to conduct inference.

Theorem 3. *Under Assumptions 1 to 7, theorems 1 and 2, and proposition 1, we have*

$$\begin{aligned}
&\sqrt{n}(\hat{\theta}^{*b}_{j|k} - \hat{\theta}^{b}_{j|k}) \xrightarrow{p} \mathbb{Z}_{j|k} \\
&\sqrt{n}(\hat{\beta}^{*}_{j|k} - \hat{\beta}_{j|k}) \xrightarrow{p} \mathbb{B}_{j|k} \\
&\sqrt{n}(\widehat{APE}^{*}_{j|k} - \widehat{APE}_{j|k}) \xrightarrow{p} N(0, V_{APE})
\end{aligned}$$

conditional on data in probability which means the bootstrap law is consistent for the limit law as n goes to infinity.

Based on Assumptions 3 to 5, we can show that all functions mentioned in the covariance functions of Theorems 1 and 2 and proposition 1 are Donsker classes by referring to Example 19.6, 19.9 and 19.20 in Van der Vaart (2000). Then by Theorem 3.6.13 in Van Der Vaart and Wellner (1996), the results can be proved. The result of Theorem 3 shows that the exchangeable bootstrap is valid for estimating the limit distributions of the vector-valued and function-valued estimators and thus provides theoretical foundation for us to derive the simultaneous or uniform confidence bands.

Next, we propose simultaneous sup-t confidence bands. Sup-t confidence bands are the Cartesian product of confidence intervals that additionally satisfy that all intervals cover the true parameter vector no less than a specific probability. They can be obtained by choosing the critical value which determine the simultaneous coverage probability; i.e., they are similar to standard, pointwise confidence intervals except a different critical value is chosen. For example, a 95% uniform confidence band would involve choosing a somewhat larger critical value than 1.96. An attractive feature of sup-t bands is they are not conservative (see Montiel Olea and Plagborg-Møller (2019) for additional discussion of these types of confidence bands).

We define confidence bands for the counterfactual transition matrix of the form:

$$\hat{B}_{j|k}(c) := [\hat{\theta}_{j|k} - \hat{\sigma}_{j|k}c\ ,\ \hat{\theta}_{j|k} + \hat{\sigma}_{j|k}c] \times \cdots \times [\hat{\theta}_{j|k} - \hat{\sigma}_{j|k}c, \hat{\theta}_{j|k} + \hat{\sigma}_{j|k}c]$$

where c is the critical value and $\hat{\sigma}_{j|k}$ is the pointwise standard error.

Next, we propose an algorithm to compute the sup-t band. We focus on transition matrices, but an analogous version of the same algorithm can also be applied to other parameters of interest. Before going into details, we first define some notations. Define the transition matrices for black and white families as M_b and M_w.

Set M_1 as the transition matrix with each element equal to 0.25 and M_c as the counterfactual transit on matrix. Let $\Theta_1 = vec(M_1)$,$\Theta_c = vec(M_c)$, $\Theta_b = vec(M_b)$ and $\Theta_w = vec(M_w)$. Also θ_1, θ_c, $\theta^b j|k$ and $\theta^w_{j|k}$ represent the element of each transition matrix. We are interested in testing and constructing the uniform confidence intervals for $\Theta_b - \Theta_1$, $\Theta_w - \Theta_1$, $\Theta_b - \Theta_c$ and $\Theta_w - \Theta_c$. It immediately follows from the previous results that

$$\sqrt{n}\Big(\hat{\Theta}_b - \Theta_1\Big) \xrightarrow{d} N(0, V_{b1}); \qquad \sqrt{n}\Big((\hat{\Theta}_b - \hat{\Theta}_c) - (\Theta_b - \Theta_c)\Big) \xrightarrow{d} N(0, V_{bc})$$

$$\sqrt{n}\Big(\hat{\Theta}_w - \Theta_1\Big) \xrightarrow{d} N(0, V_{w1}); \qquad \sqrt{n}\Big((\hat{\Theta}_w - \hat{\Theta}_c) - (\Theta_w - \Theta_c)\Big) \xrightarrow{d} N(0, V_{wc})$$

where, for the equations on the left, the result holds under the null that children are equally likely to move to all quartiles of the income distribution. Moreover, we can use the following algorithm to conduct uniform inference.

Algorithm 1. 1. Draw B samples of size N from the original sample.

2. For $b = 1 : B$, calculate the transition matrix $\hat{\Theta}^b$.

3. Calculate the empirical standard deviations $\hat{\sigma}_{j|k}$ for each cell of the transition matrix

4. For $b = 1 : B$, calculate the critical value for each sample.
$H_b = \max_{j,k} \frac{|\hat{\theta}^b_{j|k} - \hat{\theta}_1|}{\hat{\sigma}_{j|k}}$ or $\max_{i,j} \frac{|\hat{\theta}^b_{j|k} - \hat{\theta}_c|}{\hat{\sigma}_{ij}}$.

5. $\hat{q}_{1-\alpha}$ be the $1 - \alpha$ quantile of the distribution $H_1, ..., H_B$.

6. The uniform confidence intervals are shown as follows:

$$\hat{B} = \times_{j,k} [\hat{\theta}^b_{j|k} - \hat{q}_{1-\alpha}, \hat{\theta}^b_{j|k} + \hat{q}_{1-\alpha}]$$

Proposition 3 in Appendix B proves the validity of Algorithm 1.

5 Application

We use the approach discussed above to study intergenerational mobility, and, in particular, differences in transition matrices between black and white families.

5.1 Data

The data that we use comes from the National Longitudinal Study of Youth 1997 (NLSY97). The NLSY97 is a longitudinal survey that began in 1997 with a nationwide representative sample of 8,984 individual born during the years 1980 through 1984 and who were between 12 and 16 as on December 31, 1996. The first interview was conducted in 1997, and, importantly, in the first survey, data was collected about parents' income. The sample has been re-interviewed annually from 1997 to 2011 and biannually thereafter.

To construct the dataset used in the paper, we first limit the sample to children whose reported race is either black or white. For parents' income, we use reported household income for 1996. For child's income, we use their household income in 2015 (this corresponds to children being between 30 and 34 in our sample). The covariates that we use are mother's education, mother's age when child was born, age of child, child's sex, and Armed Services Vocational Aptitude Battery (ASVAB) score. We drop all observations who are missing any of the income or covariate data. This results in a dataset with 2,757 pairs of child's and parents' income along with covariates. Of these, there are 864 black families and 1,893 white families.

Summary statistics are provided in Table 1. These summary statistics are provided separately by family race and parents' income quartile.[3] From the summary statistics, there are some patterns that are immediately evident. First, parents' income for black families is noticeably more concentrated in lower quartiles. For example, 46% of black families have parents' income in the first quartile of the income distribution while only 15% of white families have parents' income in the first quartile. On the other hand, 10% of black families have parents' income in the top quartile while 32% of white families have income in the top quartile. In terms of other covariates, for both black and white families, parents' tend to be older and more educated for parents in higher income quartiles. And there are not big differences between black and white families in the same quartile in terms of parents' age or education (parents in black families tend to be slightly less educated compared to white families from the same quartile). ASVAB percentile score sharply increases across parents' income quartiles for both black and white families. More noticeably though, there are big differences in ASVAB score for black and white families within the same quartile.[4]

Next, for both white and black families, child's income is increasing in parents' income quartile; i.e., children from higher income families tend to have higher income than children from lower income families. Notably, within particular parents' income quartile, there are also large differences in child's income for black families and white families. Across all quartiles of parents' income, black children have substantially lower income than white children. For example, children from white families in the first quartile of parents' income have higher income than children from black families in the third quartile of parents' income. And children from black families in the fourth quartile of parents' income have only slightly higher income than

[3]To be specific, the first quartile summary statistics come from families whose parents' income was below the 25th percentile (of the overall distribution), thee second quartile summary statistics come from families whose parents' income was between the 25th and 50th percentiles, etc.

[4]It is not clear whether or not ASVAB should be included as a covariate in our application as the ASVAB test is taken as a teenager and parents' income could therefore affect ASVAB scores directly. That said, ASVAB is pre-labor market and therefore is perhaps useful for understanding the timing of differences in the transition matrix of black and white families. For most of the results in the application, we report two sets of results where one includes ASVAB and the other does not.

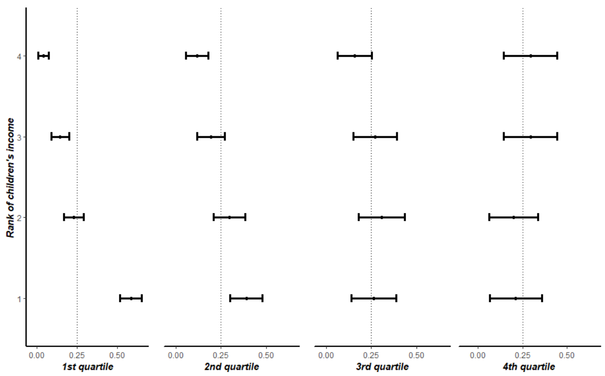
Table 1: Summary Statistics

		Parents' quartile			
		1st	2nd	3rd	4th
Black Families	Male	0.430 (0.496)	0.463 (0.500)	0.451 (0.499)	0.471 (0.502)
	Child's Age	14.16 (1.46)	14.28 (1.47)	14.53 (1.46)	14.42 (1.50)
	ASVAB	21.59 (18.65)	32.99 (24.81)	39.66 (23.37)	48.60 (27.73)
	Mother's Age	23.36 (5.69)	23.60 (4.86)	24.83 (4.43)	26.69 (5.46)
	Mother's Educ	11.70 (1.78)	12.54 (1.99)	13.35 (1.83)	14.05 (2.34)
	Urban	0.81 (0.39)	0.76 (0.43)	0.80 (0.40)	0.93 (0.26)
	Parents' Inc	10.7 (6.9)	32.3 (5.3)	51.8 (6.3)	92.4 (24.0)
	Child's Inc	35.7 (37.4)	54.5 (52.6)	68.0 (60.8)	83.8 (63.0)
	No. Families	400	246	133	85
White Families	Male	0.500 (0.501)	0.509 (0.500)	0.514 (0.500)	0.510 (0.500)
	Child's Age	14.14 (1.44)	14.26 (1.50)	14.24 (1.43)	14.24 (1.45)
	ASVAB	46.65 (27.18)	54.92 (27.77)	60.51 (26.87)	67.43 (24.01)
	Mother's Age	24.38 (5.42)	24.93 (4.95)	26.09 (4.75)	27.94 (4.46)
	Mother's Educ	12.41 (2.36)	12.93 (2.11)	13.34 (2.24)	14.57 (2.45)
	Urban	0.61 (0.49)	0.62 (0.49)	0.65 (0.48)	0.67 (0.47)
	Parents' Inc	14.5 (6.4)	34.1 (5.4)	53.3 (6.6)	106.3 (50.3)
	Child's Inc	66.4 (62.9)	78.2 (58.4)	88.8 (64.6)	110.3 (79.6)
	No. Families	290	444	555	604

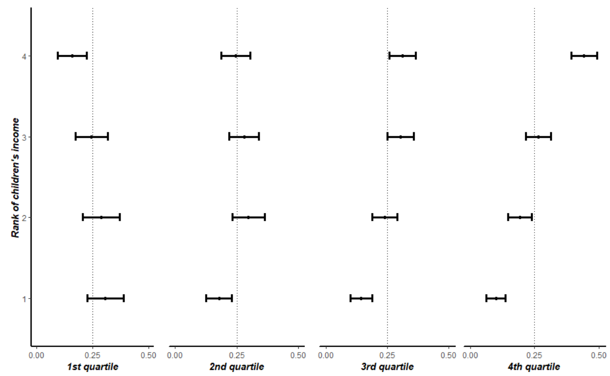
Notes: Summary statistics by race and parents' income quartile. ASVAB is the percentile score on the ASVAB test. Mother's age is their age in 1997 at the original survey. Parents' and child's income are in thousands of dollars. Standard errors are reported in parentheses.

children from white families in the second quartile of parents' income. This immediately suggests that intergenerational mobility measures are likely to be quite different for black and white families. Finally, it is worth noting that, within each quartile, parents' income tends to be lower for black families than for white families. Bhattacharya and Mazumder (2011) point out this sort of phenomenon using different data and point out that this is a potential limitation of transition matrices. In particular, it means that, on average, black children have to increase their income relative to their parents' by more to move to the next quartile than children from white families. These differences are most pronounced in the first and fourth quartiles.

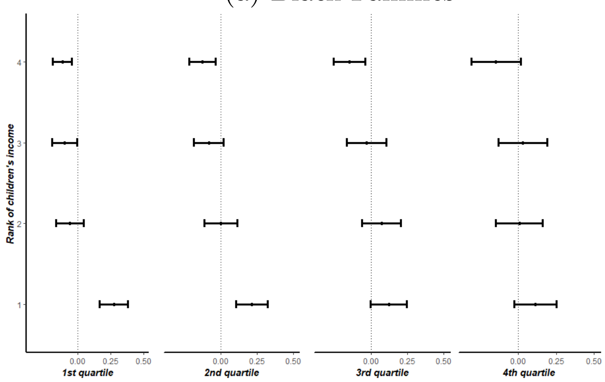
Figure 1: Unconditional Transition Matrices



(a) Black Families



(b) White Families



(c) Difference

Notes: Estimated unconditional transition matrix for black families (Panel a), white families (Panel b), and their difference (Panel c). Each column in the figure corresponds to the quartile of parents' income, and the rows provide the fraction of children whose income falls into a particular quartile conditional on their parents' income quartile. Panel (c) provides the estimate of each cell in the transition matrix for black families minus the corresponding cell in the transition matrix for white families. The estimates include a 95% uniform confidence interval constructed using the bootstrap (as discussed in the text) with 500 bootstrap iterations.

5.2 Results

Unconditional Transition Matrices

To start with, we report unconditional transition matrices separately for black and white families. Throughout this section, we report transition matrices in figures rather than in tables though we provide the tables in Appendix A. Reporting figures is especially useful for reporting uniform confidence intervals that are robust to the multiple testing involved in estimating the transition matrix.

These results are reported in Figure 1. Panel (a) of Figure 1 reports the transition matrix for black families. In the figure, the null hypothesis is that, for all quartiles of parents' income, children are equally likely to be in any quartile of the income distribution (i.e., this sort of null is consistent with parents' income having no effect on child's income). Immediately, it is clear from the figure that this null is rejected. For black families, children with parents' whose income is in the first quartile are more likely to remain in the first quartile (we estimate that 58% remain in the first quartile) and substantially less likely to move to the the third or fourth quartile of the income distribution (15% to the third quartile and 4% to the fourth quartile). Similarly, children from the second quartile are noticeably less likely to move to the upper quartile of the income distribution (only 12%) while we estimate that 39% move to the lowest quartile of the income distribution. On the other hand, for children from families whose income is in the top quartile, we do not reject that they are equally like to be in any quartile of the income distribution (though the point estimates are in line with children from higher income families being at least somewhat more likely to be in higher quartiles).

Unconditional transition matrices for white families are provided in Panel (b) of Figure 1. Children in white families with parents' income in the first quartile are most likely to remain in the first quartile, but the gradient is not nearly so steep as for black families. We estimate the 31% of children in white families with parents' income in the first quartile remain in the first quartile while 10% move to the fourth quartile. Perhaps the most interesting pattern in the transition matrix for white families is that the most common quartile for child's income is equal to the quartile of parents' income – and this holds for all quartiles of parents' income. And, in addition, for children in white families whose parents' income is in the top quartile, 44% remain in the top quartile and another 27% move to the third quartile. A small, but notable, fraction (10%) move to the first quartile.

Panel (c) of Figure 1 plots the difference between the transition matrices of black families and white families along with uniform confidence intervals that immediately allow for a test of whether or not the two transition matrices are equal to each other. It is clear from the figure that the transition matrices of black families and white families are statistically different from each other. The biggest differences occur for children moving to the lowest quartile and highest quartile of the income distribution. Across all quartiles of parents' income, children from white families are more likely to move into the fourth quartile of the income distribution; across parents' income quartile children from white families are 11-15% more likely to move to the top quartile of the income distribution (the results are only marginally statistically different in the fourth quartile though the point estimates are still relatively large). On the other hand, across all quartiles of parents' income, children from black families are more likely to move to the first quartile of the income distribution. These gaps are most pronounced for children from families with parents' income in the lowest quartile of the income distribution where the gap is 28%; the gap is smaller for higher quartiles of parents' income (e.g., 11% for children from families in the fourth quartile of parents' income) but still relatively large.

Counterfactual Transition Matrices

Next, we report counterfactual transition matrices for black families if they had the same distribution of characteristics as white families *in the same quartile of parents' income*. In this part, we use the semiparametric single-index binary response model to estimate the transition matrices of black and white families. Estimates of these counterfactual transition matrices are provided in Figure 2 (the corresponding transition matrix in table form is provided in Table 6 in Appendix A) and comparisons with the observed transition matrices of black and white families are provided in Table 2.

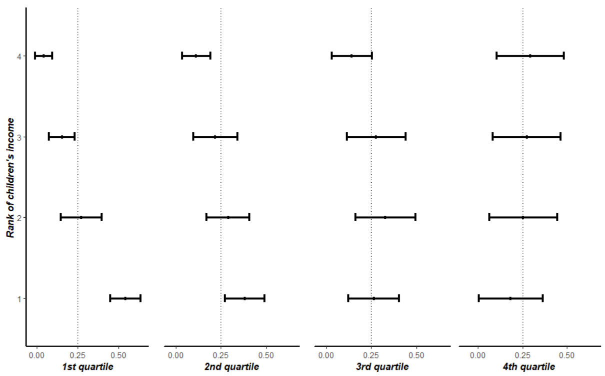
Compared to the actual transition matrix in Figure 1, adjusting for differences in covariates of black and white families somewhat explains the strong intergenerational income persistence for black families. Starting with the set of covariates that does not include ASVAB, the main differences between the counterfactual and actual transition matrices for black families is for families with parents' income in the first quartile.[5] For these families, equalizing the distribution of covariates for black and white families would move about 4% of children from the lowest income quartile to the second quartile. The other difference is for children from black families with parents' income in the top quartile who appear to be somewhat less likely to move to the lowest income quartile. The difference between the counterfactual and actual transition matrices is more pronounced when accounting for ASVAB scores (in addition to the other covariates). In this case, accounting for differences in the distribution of covariates between black and white families, 16% less children from black families with parents' income in the first quartile stay in the first income quartile relative to the observed transition matrix.

Next, it is interesting to compare the counterfactual transition matrix to the observed transition matrix for white families.[6] Without including ASVAB as a covariate, the differences

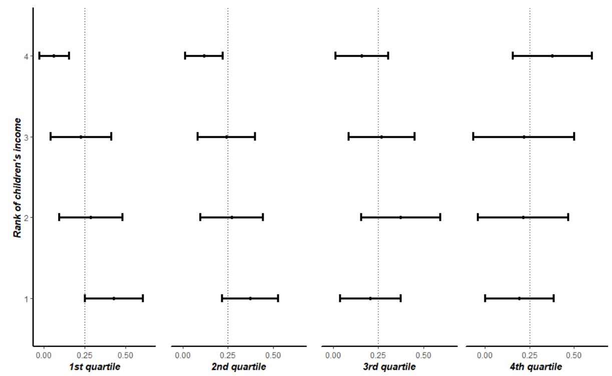
[5] Comparisons between the counterfactual transition matrix and actual transition matrix for black families is similar to a "composition effect" in the literature on decompositions (e.g., Fortin, Lemieux, and Firpo (2011)) as it arises due to differences in the distribution of observed characteristics across groups.

[6] This sort of comparison is similar to "structure effects" in the decomposition literature. Under certain extra conditions, these effects could be treatment effects or due to discrimination. We do not make these sorts of assumptions

Figure 2: Counterfactual Transition Matrices



(a) No ASVAB



(b) ASVAB

Notes: Counterfactual transition matrices for black families if their characteristics are adjusted to follow the same distribution as characteristics of white families with parents' income in the same quartile. Panel (a) does not include ASVAB as a covariate while Panel (b) includes ASVAB. The figure contains 95% uniform confidence intervals computed using the procedure discussed in the text based on the bootstrap with 500 iterations.

between the counterfactual transition matrix for black families and the actual transition matrix for white families is broadly similar to the difference between the two actual transition matrices (mainly the difference between the probability of children staying in the first quartile is somewhat mitigated, but there is still a large gap between black and white families). Including ASVAB as a covariate leads to smaller (though still large differences); in particular, by far the biggest difference is that, accounting for differences in ASVAB, 12% less children from white families with parents' income in the first quartile stay in the first quartile of the income distribution (without adjusting for covariates, 28% less children from white families stay in the first quartile). Besides this notable case, other differences are much smaller.

Overall, the results in this section suggest that adjusting for differences in the distribution of covariates for black and white families does somewhat explain the differences between the actual transition matrices of the two group. However, even after accounting for these differences, there remain large, unexplained differences between black and white families in terms of their intergenerational mobility.

here, but, rather, interpret these differences as differences between black and white families that are not explained by the covariates that are included in our application.

Table 2: Comparisons between counterfactual and actual transition matrices

Parents' Quartile	Child's Quartile 1	2	3	4
1	-0.045^{***}	0.041^{***}	-0.007	-0.002
	(0.014)	(0.013)	(0.008)	(0.005)
2	-0.009	0.008	0.023^{*}	-0.005
	(0.011)	(0.017)	(0.013)	(0.010)
3	0.002	-0.017	0.003	-0.019
	(0.019)	(0.021)	(0.024)	(0.021)
4	-0.029	0.053^{*}	-0.022	-0.002
	(0.021)	(0.029)	(0.035)	(0.028)

(a) Difference between counterfactual and actual transition matrix for black families (w/o ASVAB)

Parents' Quartile	Child's Quartile 1	2	3	4
1	0.230^{***}	-0.019	-0.093^{***}	-0.119^{***}
	(0.038)	(0.033)	(0.031)	(0.023)
2	0.203^{***}	-0.009	-0.061	-0.133^{***}
	(0.035)	(0.047)	(0.039)	(0.029)
3	0.117^{***}	0.086^{*}	-0.030	-0.172^{***}
	(0.048)	(0.048)	(0.046)	(0.036)
4	0.083	0.059	0.007	-0.150^{***}
	(0.051)	(0.049)	(0.053)	(0.057)

(b) Difference between counterfactual transition matrix of black families (w/o ASVAB) and actual transition matrix of white families

Parents' Quartile	Child's Quartile 1	2	3	4
1	-0.156^{***}	0.055	0.081	0.020
	(0.053)	(0.071)	(0.054)	(0.028)
2	-0.018	-0.027	0.046	-0.001
	(0.045)	(0.052)	(0.050)	(0.029)
3	-0.059	0.064	-0.004	-0.001
	(0.043)	(0.065)	(0.057)	(0.034)
4	-0.020	0.013	-0.077	0.083
	(0.052)	(0.082)	(0.094)	(0.066)

(c) Difference between counterfactual and actual transition matrix for black families (w/ ASVAB)

Parents' Quartile	Child's Quartile 1	2	3	4
1	0.119^{***}	-0.004	-0.019	-0.096^{***}
	(0.059)	(0.077)	(0.063)	(0.037)
2	0.194^{***}	-0.028	-0.038	-0.128^{***}
	(0.056)	(0.070)	(0.059)	(0.040)
3	0.060	0.133^{*}	-0.038	-0.155^{***}
	(0.054)	(0.071)	(0.068)	(0.049)
4	0.093	0.020	-0.048	-0.065
	(0.068)	(0.085)	(0.106)	(0.083)

(d) Difference between counterfactual transition matrix for black families (w/ ASVAB) and actual transition matrix of white families

Notes: Differences between the counterfactual transition matrix of black families if their characteristics where adjusted to be the same as white families with parents' income in the same quartile. Each cell in the table is the difference between that cell in the counterfactual transition matrix minus the corresponding cell in the actual transition matrix. The top row does not include ASVAB as a covariate; the bottom row includes ASVAB as a covariate. Pointwise standard errors are reported that were calculated using the bootstrap with 500 iterations. * indicates statistically significant at the 10% level, ** indicates statistically significant at the 5% level, and *** indicates statistically significant at the 1% level.

Average Partial Effects

In light of the results in the previous section, an immediate follow-up question is: which covariates contribute to the differences between the counterfactual and observed transition matrices? The role of particular covariates in explaining gaps between groups (i.e., a detailed decomposition) is often complicated in nonlinear models (see the discussion in Fortin, Lemieux, and Firpo

Table 3: Average Partial Effects for Families in the First Quartile of Parents' Income

		Child's quartile			
		1st	2nd	3rd	4th
Black Family	Age	1.0602** (0.4791)	0.0997 (0.7896)	-1.8895 (1.1974)	0.7297 (0.8450)
	Male	−0.2041** (0.1039)	0.0995 (0.1082)	0.1059 (0.0863)	-0.0014 (0.0479)
	Mother's Age	-0.0031 (0.0057)	0.0030 (0.0076)	0.0001 (0.0072)	0.0001 (0.0043)
	Mother's Educ	−0.0288*(0.0162)	0.0146 (0.0223)	0.0100 (0.0199)	0.0043 (0.0105)
	Urban	0.0717 (0.0766)	-0.1043 (0.0711)	0.0331 (0.0623)	-0.0006 (0.0422)
	ASVAB	−0.0080*** (0.0030)	0.0050 (0.0038)	0.0022 (0.0037)	0.0008 (0.0019)
White Family	Age	0.4798 (0.5391)	0.1521 (0.7958)	-0.0968 (0.8224)	-0.5350 (0.6537)
	Male	−0.2601*** (0.0878)	0.0842 (0.0882)	0.1233 (0.0815)	0.0526 (0.0570)
	Mother's Age	-0.0104 (0.0075)	0.0047 (0.0094)	0.0019 (0.0088)	0.0038 (0.0046)
	Mother's Educ	0.0177 (0.0178)	-0.0077 (0.0219)	0.0046 (0.0179)	-0.0146 (0.0121)
	Urban	−0.1981*** (0.0777)	-0.0208 (0.0728)	0.0582 (0.0678)	0.1606*** (0.0609)
	ASVAB	-0.0015 (0.0024)	-0.0015 (0.0031)	0.0016 (0.0024)	0.0015 (0.0019)

Notes: Average partial effects for each covariate (as discussed in text) for families with parents' income in the first quartile of the income distribution and separately by race. Pointwise standard errors are reported in parentheses that are computed using the bootstrap with 500 iterations. * indicates statistically significant at the 10% level, ** indicates statistically significant at the 5% level, and *** indicates statistically significant at the 1% level.

(2011) and Rothe (2015)). Instead of proceeding that way, we report average partial effects of each covariate. These are the average partial effect of a particular covariate on the probability of a child's income being in a particular quartile (conditional on parents' income being in a particular quartile).

Table 3 provides average partial effects estimates for families with parents' income in the first quartile. Tables 7 to 9 (all of which are available in Appendix A) provide corresponding APEs for families with parents' income in the second, third, and fourth quartiles, respectively. There are some notable patterns as well as differences between black and white families. To start with, ASVAB seems to only have an effect for black families. For example, we estimate that, for a child from a black family with parents' income in the first quartile, moving from the 25th percentile to the 75th percentile of ASVAB score would decrease the probability of remaining in the first quartile by 40 percentage points – this is a very large effect. Similarly, parents' education has a big effect for black families (though not for white families); we estimate that increasing mother's education by four years would decrease the probability of a child remaining in the first quartile by 12 percentage points. We find large differences in the effect of living in an urban location for black and white families; children from urban white families are substantially less likely to remain in the first quartile while children from urban black families are more likely to remain in the first quartile (though this estimate is not statistically significant). For other quartiles of parents' income, We generally estimate smaller and not statistically significant effects of covariates (particularly mother's education and ASVAB) on the quartile of child's income. This is related to our results on counterfactual transition matrices where the biggest differences due to adjusting for differences in the distribution of covariates across races relative to the actual transition matrix of black families occurred for families with parents' income in the first quartile.

6 Conclusion

In this paper, we have developed new methods to study conditional transition matrices, counterfactual transition matrices, and transition matrix partial effects. The key challenge for all of these is to estimate cells in a conditional transition matrix. Estimating these conditional transition matrices is closely related to estimating conditional distributions, and, in order to estimate these conditional distributions, we proposed new semiparametric distribution regression estimators. Relative to parametric distribution regression estimators, the main advantage of these estimators is that they do not require researchers to specify a link function; relative to nonparametric estimators, our approach does not suffer from the curse of dimensionality. These first-step estimators may be of independent interest and could be useful in a wide variety of applications where a researcher would like to estimate a conditional distribution but does not know the correct link function. More generally, counterfactual transition matrices are particularly useful for studying intergenerational mobility but are also relevant in any application where researchers are interested in (i) understanding the dependence between two random variables while (ii) also understanding the contribution of other covariates to this dependence.

We applied our approach to study differences in transition matrices between black families and white families. There, we documented large differences in unconditional transition matrices for black and white families, but there are also notable differences in observed characteristics such as parents' education. We found that adjusting for these differences across races did explain some of the differences in each group's actual transition matrix, but that there were still large unexplained differences.

References

[1] Abrevaya, Jason, Yu-Chin Hsu, and Robert P Lieli. "Estimating conditional average treatment effects". *Journal of Business & Economic Statistics* 33.4 (2015), pp. 485–505.

[2] Ahn, Hyungtaik. "Semiparametric estimation of a single-index model with nonparametrically generated regressors". *Econometric Theory* (1997), pp. 3–31.

[3] Andrews, Donald WK. "Empirical process methods in econometrics". *Handbook of econometrics* 4 (1994), pp. 2247–2294.

[4] Bai, Jushan and Peng Wang. "Conditional Markov chain and its application in economic time series analysis". *Journal of applied econometrics* 26.5 (2011), pp. 715–734.

[5] 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.

[6] 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.

[7] Blundell, Richard W. and James L. Powell. "Endogeneity in semiparametric binary response models". *The Review of Economic Studies* 71.3 (2004), pp. 655–679.

[8] Boreiko, Dmitri, Serguei Kaniovski, Yuri Kaniovski, and Georg Pflug. "Traces of business cycles in credit-rating migrations". *Plos one* 12.4 (2017), e0175911.

[9] Callaway, Brantly and Weige Huang. "Distributional effects of a continuous treatment with an application on intergenerational mobility". *Oxford Bulletin of Economics and Statistics* 82.4 (2020), pp. 808–842.

[10] Chen, Xiaohong, Oliver Linton, and Ingrid Van Keilegom. "Estimation of semiparametric models when the criterion function is not smooth". *Econometrica* 71.5 (2003), pp. 1591–1608.

[11] Chernozhukov, V. and V. Semenova. "Simultaneous inference for Best Linear Predictor of the Conditional Average Treatment Effect and other structural functions". *arXiv: Methodology* (2018).

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

[13] Chernozhukov, Victor, Iván Fernández-Val, and Ye Luo. "The sorted effects method: discovering heterogeneous effects beyond their averages". *Econometrica* 86.6 (2018), pp. 1911–1938.

[14] Davis, Jonathan and Bhashkar Mazumder. "The decline in intergenerational mobility after 1980". Working Paper. 2018.

[15] Fan, Qingliang, Yu-Chin Hsu, Robert P Lieli, and Yichong Zhang. "Estimation of conditional average treatment effects with high-dimensional data". *Journal of Business & Economic Statistics* (2020), pp. 1–15.

[16] 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.

[17] Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. "Decomposition methods in economics". *Handbook of labor economics*. Vol. 4. Elsevier, 2011, pp. 1–102.

[18] Horowitz, Joel L. *Semiparametric and nonparametric methods in econometrics*. Vol. 12. Springer, 2009.

[19] Ichimura, Hidehiko. "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models". *Journal of Econometrics* 58.1-2 (1993), pp. 71–120.

[20] 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.

[21] Jones, Mr Matthew T. *Estimating Markov transition matrices using proportions data: an application to credit risk.* 5-219. International Monetary Fund, 2005.

[22] Klein, Roger and Richard Spady. "An efficient semiparametric estimator for binary response models". *Econometrica* (1993), pp. 387–421.

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

[24] Montiel Olea, José Luis and Mikkel Plagborg-Møller. "Simultaneous confidence bands: Theory, implementation, and an application to SVARs". *Journal of Applied Econometrics* 34.1 (2019), pp. 1–17.

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

[26] Richey, Jeremiah and Alicia Rosburg. "Decomposing economic mobility transition matrices". *Journal of Applied Econometrics* 33.1 (2018), pp. 91–108.

[27] Rolling, Craig A., Yuhong Yang, and Dagmar Velez. "Combining estimates of conditional treatment effects". *Econometric Theory* 35.6 (2019), pp. 1089–1110.

[28] Rothe, Christoph. "Semiparametric estimation of binary response models with endogenous regressors". *Journal of Econometrics* 153.1 (2009), pp. 51–64.

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

[30] Rothe, Christoph. "Decomposing the composition effect: the role of covariates in determining between-group differences in economic outcomes". *Journal of Business & Economic Statistics* 33.3 (2015), pp. 323–337.

[31] Sherman, Robert P. "Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators". *Ann. Statist.* 22.1 (Mar. 1994), pp. 439–459. URL: https://doi.org/10.1214/aos/1176325377.

[32] Vaart, Aad W van der, Jon A Wellner Wellner, et al. "Empirical processes indexed by estimated functions". *Asymptotics: particles, processes and inverse problems.* Institute of Mathematical Statistics, 2007, pp. 234–252.

[33] Van der Vaart, Aad W. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000.

[34] Van Der Vaart, Aad W and Jon A Wellner. "Weak convergence". *Weak convergence and empirical processes*. Springer, 1996, pp. 16–28.

A Additional Tables

Table 4: Point Estimates of Unconditional Transition Matrices

		Child's Quartile			
		1	2	3	4
Parents' Quartile	1	0.583***	0.230***	0.145***	0.042***
		(0.022)	(0.020)	(0.018)	(0.010)
	2	0.390***	0.297***	0.195***	0.118***
		(0.030)	(0.028)	(0.025)	(0.021)
	3	0.263***	0.308***	0.271***	0.158***
		(0.040)	(0.042)	(0.039)	(0.032)
	4	0.212***	0.200***	0.294***	0.294***
		(0.047)	(0.046)	(0.050)	(0.051)

(a) Transition Matrix of Black Families

		Child's Quartile			
		1	2	3	4
Parents' Quartile	1	0.307***	0.290***	0.245***	0.159***
		(0.026)	(0.027)	(0.024)	(0.021)
	2	0.178***	0.297***	0.279***	0.245***
		(0.018)	(0.022)	(0.021)	(0.020)
	3	0.144***	0.240***	0.305***	0.312***
		(0.015)	(0.017)	(0.018)	(0.018)
	4	0.099***	0.194***	0.265***	0.442***
		(0.012)	(0.016)	(0.017)	(0.017)

(b) Transition Matrix of White Families

		Child's Quartile			
		1	2	3	4
Parents' Quartile	1	0.276***	−0.060*	−0.100***	−0.116***
		(0.035)	(0.035)	(0.030)	(0.024)
	2	0.212***	-0.001	−0.084***	−0.128***
		(0.037)	(0.037)	(0.035)	(0.030)
	3	0.119***	0.069	-0.034	−0.154***
		(0.044)	(0.047)	(0.046)	(0.039)
	4	0.112**	0.006	0.029	−0.148***
		(0.049)	(0.048)	(0.054)	(0.055)

(c) Difference between Transition Matrices of Black and White Families

Notes: The table provides point estimates corresponding the transition matrix in Figure 1. The standard errors reported in the table are pointwise. * indicates statistically significant at the 10% level, ** indicates statistically significant at the 5% level, and *** indicates statistically significant at the 1% level.

Table 6: Counterfactual Transition Matrices for Black Families

	Parental quartile	Children quartile 1st	2nd	3rd	4th
Counterfactual Tran. Matrix with ASVAB	1st	0.426	0.285	0.226	0.062
		(0.0576)	(0.0630)	(0.0605)	(0.0292)
	2nd	0.372	0.269	0.241	0.117
		(0.0506)	(0.0561)	(0.0516)	(0.0336)
	3rd	0.204	0.373	0.266	0.157
		(0.0542)	(0.0713)	(0.0594)	(0.0479)
	4th	0.192	0.213	0.217	0.377
		(0.0626)	(0.0820)	(0.0922)	(0.0719)
Counterfactual Tran. Matrix without ASVAB	1st	0.537	0.271	0.152	0.040
		(0.0312)	(0.0426)	(0.0267)	(0.0176)
	2nd	0.381	0.289	0.218	0.113
		(0.0370)	(0.0403)	(0.0415)	(0.0266)
	3rd	0.262	0.325	0.274	0.139
		(0.0472)	(0.0563)	(0.0549)	(0.0377)
	4th	0.183	0.253	0.272	0.292
		(0.0605)	(0.0643)	(0.0644)	(0.0641)
Difference	1st	-0.111	0.015	0.074	0.022
		(0.0515)	(0.0711)	(0.0539)	(0.0274)
	2nd	-0.009	-0.019	0.023	0.005
		(0.0440)	(0.0539)	(0.0507)	(0.0283)
	3rd	-0.057	0.048	-0.008	0.018
		(0.0410)	(0.0630)	(0.0565)	(0.0339)
	4th	0.009	-0.039	-0.055	0.085
		(0.0553)	(0.0869)	(0.0945)	(0.0644)

Notes: The transition matrices shown above are counterfactual transition matrices of black families have they had the characteristics of white families. The top two panels correspond to Panels (a) and (b) of Figure 2. The third panel reports the difference between the two counterfactual distributions. Pointwise standard errors are reported that are calculated using the bootstrap with 500 iterations.

Table 7: Average Partial Effects for Families in the Second Quartile of Parents' Income

		Child's quartile			
		1st	2nd	3rd	4th
Black Family	Age	-0.4245 (0.6277)	1.1057 (0.8951)	-0.1225 (1.0001)	-0.5587 (0.8149)
	Male	0.0666 (0.1320)	0.0124 (0.1213)	-0.1405 (0.1106)	0.0616 (0.0899)
	Mother's Age	0.0032 (0.0081)	-0.0100 (0.0102)	0.0090 (0.0091)	-0.0022 (0.0061)
	Mothers' Educ	0.0047(0.0161)	-0.0062 (0.0242)	-0.0051 (0.0249)	0.0065 (0.0190)
	Urban	-0.1064 (0.1039)	0.1491 (0.0997)	-0.0692 (0.0955)	0.0266 (0.0614)
	ASVAB	-0.0010 (0.0033)	0.0016 (0.0045)	0.0024 (0.0033)	0.0001 (0.0022)
White Family	Age	1.0849 (0.5186)	0.1449 (0.7779)	-0.3815 (0.9884)	-0.8482 (0.6721)
	Male	-0.0597 (0.0643)	0.2211 (0.0671)	-0.0571 (0.0751)	-0.1043 (0.0610)
	Mother's Age	-0.0025 (0.0050)	0.0043 (0.0070)	-0.0024 (0.0074)	0.0043 (0.0049)
	Mother's Educ	-0.0113 (0.0110)	0.0054 (0.0205)	0.0044 (0.0246)	0.0015 (0.0134)
	Urban	-0.0512 (0.0629)	0.1141 (0.0613)	-0.0816 (0.0690)	0.0188 (0.0630)
	ASVAB	-0.00116 (0.0016)	0.00003 (0.0025)	0.00051 (0.0026)	0.00062 (0.0015)

Notes: Average partial effects for each covariate (as discussed in text) for families with parents' income in the second quartile of the income distribution and separately by race. Pointwise standard errors are reported in parentheses that are computed using the bootstrap with 500 iterations. * indicates statistically significant at the 10% level, ** indicates statistically significant at the 5% level, and *** indicates statistically significant at the 1% level.

Table 8: Average Partial Effects for Families in the Third Quartile of Parents' Income

		Child's quartile			
		1st	2nd	3rd	4th
Black Family	Age	3.1101 (0.9872)	-3.1890 (1.2443)	0.6557 (1.0705)	-0.5769 (0.8443)
	Male	0.4136 (0.1651)	-0.5071 (0.1639)	-0.0476 (0.1668)	0.1412 (0.1531)
	Mother's Age	0.0037 (0.0106)	-0.0013 (0.0164)	0.0009 (0.0152)	-0.0034 (0.0101)
	Mother's Educ	0.0038(0.0205)	-0.0069 (0.0305)	0.0123 (0.0355)	-0.0092 (0.0266)
	Urban	0.2036 (0.0879)	-0.2661 (0.1594)	-0.0574 (0.1318)	0.1198 (0.1091)
	ASVAB	-0.0037 (0.0032)	0.0030 (0.0046)	0.0006 (0.0040)	0.0001 (0.0021)
White Family	Age	1.2135 (0.5395)	0.0471 (0.7066)	-1.8452 (0.7331)	0.5846 (0.5893)
	Male	0.0356 (0.0550)	0.0064 (0.0682)	-0.0208 (0.0673)	-0.0212 (0.0713)
	Mother's Age	-0.0043 (0.0063)	0.0040 (0.0081)	-0.0003 (0.0075)	0.0006 (0.0046)
	Mother's Educ	0.0189 (0.0120)	-0.0317 (0.0176)	0.0141 (0.0175)	-0.0013 (0.0141)
	Urban	0.0209 (0.0461)	0.0165 (0.0545)	-0.0437 (0.0657)	0.0062 (0.0611)
	ASVAB	-0.0018 (0.0015)	-0.0001 (0.0028)	0.0003 (0.0030)	0.0014 (0.0019)

Notes: Average partial effects for each covariate (as discussed in text) for families with parents' income in the third quartile of the income distribution and separately by race. Pointwise standard errors are reported in parentheses that are computed using the bootstrap with 500 iterations. * indicates statistically significant at the 10% level, ** indicates statistically significant at the 5% level, and *** indicates statistically significant at the 1% level.

Table 9: Average Partial Effects for Families in the Second Quartile of Parents' Income

		Child's quartile			
		1st	2nd	3rd	4th
Black Family	Age	0.3151 (1.1227)	-0.3866 (1.1841)	0.6797 (0.8044)	-0.6082 (0.6268)
	Male	-0.6919 (0.3467)	0.3603 (0.3317)	-0.3350 (0.3458)	0.6667 (0.3281)
	Mother's Age	-0.0060 (0.0167)	0.0049 (0.0196)	-0.0003 (0.0159)	0.0014 (0.0124)
	Mother's Educ	0.0058(0.0393)	-0.0252 (0.0242)	0.0255 (0.0385)	-0.0061 (0.0217)
	Urban	-0.4650 (0.3471)	0.2833 (0.0995)	-0.0811 (0.3585)	0.2629 (0.0702)
	ASVAB	0.0001 (0.0078)	-0.0001 (0.0091)	-0.0020 (0.0080)	0.0019 (0.0063)
White Family	Age	0.6301 (0.7297)	0.6381 (1.0007)	-0.4767 (0.9865)	-0.7916 (0.6579)
	Male	-0.0165 (0.0498)	0.1069 (0.0623)	-0.1859 (0.0807)	-0.0955 (0.0684)
	Mother's Age	-0.0008 (0.0038)	0.0076 (0.0084)	-0.0018 (0.0099)	-0.0049 (0.0071)
	Mother's Educ	-0.0023 (0.0063)	-0.0113 (0.0125)	0.0118 (0.0159)	0.0017 (0.0103)
	Urban	-0.0401 (0.0455)	0.0273 (0.0495)	-0.1327 (0.0737)	0.1454 (0.0658)
	ASVAB	-0.0008 (0.0011)	-0.0002 (0.0021)	-0.0002 (0.0034)	0.0012 (0.0025)

Notes: Average partial effects for each covariate (as discussed in text) for families with parents' income in the fourth quartile of the income distribution and separately by race. Pointwise standard errors are reported in parentheses that are computed using the bootstrap with 500 iterations. * indicates statistically significant at the 10% level, ** indicates statistically significant at the 5% level, and *** indicates statistically significant at the 1% level.

B Proofs

In this section, we prove the theorems mentioned in the main text. We usually prove the weak convergence of CDF by making some high-level assumptions and then provide some primitive conditions to show that those high-level conditions can be satisfied. This process can be applied to parametric, semiparametric and nonparametric situation. Therefore, we first present a general form to briefly discuss those high level conditions and then discuss the primitive conditions for parametric and nonparametric situation. Finally, we explicitly discuss the conditions in semiparametric situation which is the main focus of this paper.

We first introduce some notations. Let $\tau = (\tau_1, \tau_0)$ where τ_1 is the quartile of children's income and τ_0 is the quartile of parents' income. Denote the conditional distribution of independents variables of group d' as $F_{d'|\tau_0}(X^T\beta) = F_0(X^T\beta) = F_0(Z)$ where $Z = X^T\beta$. Let $\hat{F}_{d'|\tau_0}(X^T\hat{\beta}) = \hat{F}_0(X^T\hat{\beta})$ and $\hat{F}_{d'|\hat{\tau}_0}(X^T\hat{\beta}) = \hat{F}_{0|\hat{\tau}_0}(X^T\hat{\beta})$. Define the distribution of children's income below τ_1 conditional on parent's income below τ_0 and covariates as $\gamma(\tau, X^T\beta) = \gamma(\tau_1, \tau_0, X^T\beta)$ for group d. We will omit the indexes of rows and columns of transition matrices for simplicity. Then we can obtain the general form as:

$$
\begin{aligned}
\sqrt{n}\left(\hat{\theta} - \theta\right) &= \sqrt{n}\left[\int \hat{\gamma}(\hat{\tau}, x^T\hat{\beta}) d\hat{F}_{0|\hat{\tau}_0}(x^T\hat{\beta}) - \gamma(\tau, x^T\beta) dF_0(x)\right] \\
&= \underbrace{\sqrt{n}\left[\int \left(\hat{\gamma}(\hat{\tau}, x^T\hat{\beta}) - \gamma(\tau, x^T\beta)\right) d(\hat{F}_{0|\hat{\tau}}(x^T\hat{\beta}) - F_0(x^T\beta)\right]}_{\Delta_1} \\
&+ \underbrace{\sqrt{n}\left[\left(\hat{\gamma}(\hat{\tau}, x^T\hat{\beta}) - \gamma(\tau, x^T\beta)\right) dF_0(x^T\beta)\right]}_{\Delta_2} \\
&+ \underbrace{\sqrt{n}\left[\int \hat{\gamma}(\hat{\tau}, x^T\hat{\beta}) d(\hat{F}_{0|\hat{\tau}}(x^T\hat{\beta}) - F_0(x^T\beta)\right]}_{\Delta_3}
\end{aligned}
$$

We want to show that the general form of $\sqrt{n}\left(\hat{\theta} - \theta\right)$ can be written as follows: If $E[\varphi] < \infty$ and $E[\varphi^2] < \infty$, then

$$
\sqrt{n}\left(\hat{\theta} - \theta\right) = \frac{1}{\sqrt{n}} \sum_i \left[\Gamma_{i1} + \Gamma_{i2}\right] + o_p(1) \xrightarrow{d} \mathbf{\Gamma}
$$

if and only if

$$\Delta_1 = o_p(1)$$
$$\Delta_2 = \frac{1}{\sqrt{n}} \sum_i \Gamma_{i1} + o_p(1) \xrightarrow{d} \mathbf{\Gamma}_1$$
$$\Delta_3 = \frac{1}{\sqrt{n}} \sum_i \Gamma_{i2} + o_p(1) \xrightarrow{d} \mathbf{\Gamma}_2$$

where $\mathbf{\Gamma}$, $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are mean zero Gaussian process.

B.1 Parametric and Nonparametric Case

In this section, we will try to show that under some conditions the weak convergence of CDF holds in parametric and nonparametric situations. Follow the notation mentioned above, Denote $\hat{\theta} = \int \Lambda(x^T\hat{\beta}, \hat{\tau}) d\hat{F}_{0|\hat{\tau}_0}(x) = \frac{1}{n}\sum_i \frac{1-D_i}{1-p}\Lambda(x_i^T\hat{\beta})$ and $\theta = \int \Lambda(x^T\beta)dF_0(x)$ and Let the true conditional distribution $\phi(\tau, x) = Pr(\tau_0, \tau_1|x)$ which is differentiable at τ. In parametric distribution regression, you can check the Appendix of Chernozhukov, Fernandez-Val, and Melly (2013) to learn the asymptotic theory of parametric distribution regression. In that paper, they estimate the conditional distribution and counterfactual distribution by setting the link function as parametric function such as logit and using the empirical distribution function as the distribution of covariates. In this part, we will briefly show the proving process of logit distribution function. We can write as follows:

$$\begin{aligned}
\sqrt{n}\left(\hat{\theta} - \theta\right) &= \int \Lambda(x^T\hat{\beta}, \hat{\tau}) d\hat{F}_{0|\hat{\tau}_0}(x) - \int \Lambda(x^T\beta)dF_0(x) \\
&= \underbrace{\sqrt{n}\left\{\int \left[\Lambda(x^T\hat{\beta}, \hat{\tau}) - \Lambda(x^T\beta)\right] dF_0(x)\right\}}_{\Delta_1} + \underbrace{\sqrt{n}\left\{\int \Lambda(x^T\beta)\left[\hat{F}_{0|\hat{\tau}_0}(x) - dF_0(x)\right]\right\}}_{\Delta_2} \\
&+ \underbrace{\sqrt{n}\left\{\int \left[\Lambda(x^T\hat{\beta}, \hat{\tau}) - \Lambda(x^T\beta)\right] d\left[\hat{F}_{0|\hat{\tau}_0}(x) - F_0(x)\right]\right\}}_{\Delta_3}
\end{aligned}$$

The first term can be written as follows:

$$\begin{aligned}
\sqrt{n}\left\{\int \left[\Lambda(x^T\hat{\beta},\hat{\tau})-\Lambda(x^T\beta)\right]dF_0(x)\right\} &= \sqrt{n}\left\{\int \left[\Lambda(x^T\hat{\beta},\hat{\tau})-\Lambda(x^T\hat{\beta},\tau)-[\Lambda(x^T\beta,\hat{\tau})-\Lambda(x^T\beta)]\right]dF_0(x)\right\}\\
&+\sqrt{n}\left\{\int \left[\Lambda(x^T\beta,\hat{\tau})-\Lambda(x^T\beta)\right]dF_0(x)\right\}\\
&+\sqrt{n}\left\{\int \left[\Lambda(x^T\hat{\beta},\tau)-\Lambda(x^T\beta)\right]dF_0(x)\right\}\\
&=\sqrt{n}\left\{\int [\Lambda_\beta(x^T\beta)(\hat{\beta}-\beta)+\Lambda_\tau(x^T\beta)(\hat{\tau}-\tau)]dF_0(x)\right\}
\end{aligned}$$

where $\Lambda_\beta(\cdot)$ is the first order derivative with respect to β and $\Lambda_\tau(\cdot)=\Lambda(\cdot)\phi_\tau(\cdot)$. The first term in the first equation will converge to zero because of the Lemma 2, since logit function is uniform bounded and also the first order derivative is also uniform bounded and thus satisfies the lipschitz condition. The behavior of the rest terms will follow from the application of delta method. The term in the last equation can be obtained by Lemma 6 and the convergence of estimator of β. We show that $\sqrt{n}(\hat{\beta}-\beta)$ converge to a normal distribution $N(0,V_l)$ by using standard maximum likelihood method. V_l is equal to the inverse of Fisher information which can be obtained by taking the second order derivative with respect to β of likelihood function. For example, we can obtained the log-likelihood function as follows:

$$\begin{aligned}
\hat{\beta} &= \operatorname*{argmin}_\beta \mathcal{L}\\
&= \operatorname*{argmin}_\beta \sum_i y_i log(\Lambda(x_i^T\beta))+(1-y_i)(1-\Lambda(x_i^T\beta))
\end{aligned}$$

Taking the first order condition and choosing the $\hat{\beta}$ as the optimal value, then we can get

$$0=\sum_i\left[\frac{y_i\Lambda^1(x_i^T\hat{\beta})}{\Lambda(x_i^T\hat{\beta})}-\frac{(1-y_i)\Lambda^1(x_i^T\hat{\beta})}{1-\Lambda(x_i^T\hat{\beta})}\right]x_i^T$$

Using Taylor expansion, we can get

$$
\begin{aligned}
0 &= \sum_i \left[\frac{y_i \Lambda'(x_i^T \hat{\beta})}{\Lambda(x_i^T \hat{\beta})} - \frac{(1-y_i)\Lambda'(x_i^T \hat{\beta})}{1-\Lambda(x_i^T \hat{\beta})} \right] x_i^T \\
&= \sum_i \underbrace{\left[\frac{y_i \Lambda'(x_i^T \beta)}{\Lambda(x_i^T \beta)} - \frac{(1-y_i)\Lambda'(x_i^T \beta)}{1-\Lambda(x_i^T \beta)} \right] x_i^T}_{\ell_i'} \\
&+ \sum_i \underbrace{\left\{ \left[-\frac{y_i \Lambda'(x_i^T \beta)}{\Lambda(x_i^T \beta)} - \frac{(1-y_i)\Lambda'(x_i^T \beta)}{1-\Lambda(x_i^T \beta)} \right] \Lambda' + \left[\frac{y_i}{\Lambda(x_i^T \beta)} - \frac{1-y_i}{1-\Lambda(x_i^T \beta)} \right] \Lambda'' \right\} x_i x_i^T}_{\ell_i''} (\hat{\beta} - \beta)
\end{aligned}
$$

where $\ell_i'' = -\frac{exp(z_i)}{(1+exp(z_i))^2} x_i x_i^T$ and $z_i = x_i^T \beta$. Then

$$
\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &= \frac{\sqrt{n}\left(\frac{1}{n}\sum_i \ell_i' - E(\ell_i') \right)}{\frac{1}{n}\sum_i \ell_i''} \\
&\xrightarrow{d} N(0, I(\beta)^{-1})
\end{aligned}
$$

where $E(\ell_i') = 0$ and $I(\beta) = -E(\ell_i'') = E(\frac{exp(x_i^T\beta)}{(1+exp(x_i^T\beta))^2} x_i x_i^T)$. Then the first term can be obtained by plugging in the equation of $\sqrt{n}(\hat{\beta} - \beta)$.

For Δ_2, we can get:

$$
\begin{aligned}
\Delta_2 =& \sqrt{n}\left\{ \int \Lambda(x^T\beta) d\left[\hat{F}_{0|\hat{\tau}_0}(x) - F_{0|\hat{\tau}_0}(x) - [\hat{F}_{0|\tau_0}(x) - F_0(x)] \right] \right\} \\
&+ \sqrt{n}\left\{ \int \Lambda(x^T\beta) d\left[\hat{F}_{0|\tau_0}(x) - F_0(x) \right] \right\} + \sqrt{n}\left\{ \int \Lambda(x^T\beta) d\left[F_{0|\hat{\tau}_0}(x) - F_0(x) \right] \right\}
\end{aligned}
$$

The first term will converge to zero by the stochastic equicontinuity. Since $\mathcal{F} = \{1(X \leq x)$, for $x \in \mathcal{X}$, and indexed by $\tilde{\tau}_0 \in \mathcal{T}\}$ is a VC-class and thus belongs to the type-I class in Andrews (1994) which is the sufficient condition for stochastic equicontinuity. The second term will converge to a mean zero Gaussian process by functional central limit theorem since indicator functions belongs to Donsker class. The third term satisfies weak convergence by taking Taylor expansion. $F_{0|\hat{\tau}_0}(x) - F_0(x) = \int \phi^p_{\tau_0}(\tau_0, x) f(x)/f(\tau_0) dx(\tau_0 - \tau_0)$ where $\phi^p(\tau_0, x) = Pr(\tau_0|x)$ is differentiable with respect to τ_0.

Δ_3 will be equal to $o_p(1)$ by the results we get from Δ_1 and Δ_2. We can see that distribution estimated by logit distribution regression does convergence to a normal distribution with $\sqrt{n}$ convergence rate.

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \xrightarrow{d} N(0, V_{logit})$$

where $V_{logit} = E[E[\Lambda_\beta](\hat{\beta}-\beta) + E[\Lambda_{\tau_0} + \Lambda\frac{\phi^p_{\tau_0}}{\phi^p}](\hat{\tau}_0 - \tau_0) + E[\Lambda_{\tau_1}](\hat{\tau}_1 - \tau_1) + \frac{1}{P^{d'}(\tau_0)}(\Lambda(x^T\beta) - E(\Lambda(x^T\beta))]^2$.

For the nonparametric case, please refer to T_1 term in the Semiparametric case below.

B.2 Semiparametric Case

For the trimming set A_x, Horowitz (2009) points out that Klein and Spady (1993) use complicated trimming procedures to accomplish this without artificially restricting X to a fixed set A_x on which $g(X^T\beta)$ is bounded away from 0 and 1. They find, however, that trimming has little effect on the numerical performance of the resulting estimator. Therefore, in practice, they suggest following the simpler approach of using only observations which satisfy $X \in A_x$. Here we let A_x to be a fixed set in order to simplify the proof and set $\alpha_i = 1\{x_i \in A_x\}$. Since the assumptions in main text imply that the support of covariates is equal to A_x, we will omit the indicator functions when we show the asymptotic theory for simplicity without loss of generality.

We first introduce some new notations. We replace G in the main text with g. Let $\tau = (\tau^c, \tau^p) = (\tau_1, \tau_0)$ where τ_1 is the quartile of children's income and τ_0 is the quartile of parents' income. Denote the conditional distribution of independents variables of group d' as $F_{d'|\tau_0}(X^T\beta) = F_0(X^T\beta) = F_0(Z)$ where $Z = X^T\beta$. Let $\hat{F}_{d'|\tau_0}(X^T\hat{\beta}) = \hat{F}_0(X^T\hat{\beta})$ and $\hat{F}_{d'|\hat{\tau}_0}(X^T\hat{\beta}) = \hat{F}_{0|\hat{\tau}_0}(X^T\hat{\beta})$. Define the distribution of children's income below τ_1 conditional on parent's income below τ_0 and covariates as $g(\tau, X^T\beta) = g(\tau_1, \tau_0, X^T\beta)$. We will omit the indexes of rows and columns of transition matrices for simplicity. Then we can obtain the general form as:

$$\sqrt{n}\left(\hat{\theta}-\theta\right) = \sqrt{n}\left[\int \hat{g}(\hat{\tau}, x^T\hat{\beta})d\hat{F}_{0|\hat{\tau}_0}(x^T\hat{\beta}) - g(\tau, x^T\beta)dF_0(x)\right]$$
$$= \underbrace{\sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - \hat{g}(\tau, x^T\hat{\beta})\right) d\hat{F}_0(x^T\hat{\beta})\right]}_{\Delta_1} + \underbrace{\sqrt{n}\left[\int \hat{g}(\tau, x^T\hat{\beta})d\hat{F}_0(x^T\hat{\beta}) - \int g(\tau, x^T\beta)dF_0(x)\right]}_{\Delta_2}$$
$$+ \underbrace{\sqrt{n}\left[\int \hat{g}(\hat{\tau}, x^T\hat{\beta})d(\hat{F}_{0|\hat{\tau}}(x^T\hat{\beta}) - \hat{F}_0(x^T\hat{\beta})\right]}_{\Delta_3}$$

Proof of Theorem 1. Beginning with the third term, we have

$$\begin{aligned}
\Delta_3 &= \sqrt{n}\left[\int \hat{g}(\hat{\tau}, x^T\hat{\beta})d(\hat{F}_{0|\hat{\tau}}(x^T\hat{\beta}) - \hat{F}_0(x^T\hat{\beta}))\right] \\
&= \sqrt{n}\left[\int \hat{g}(\hat{\tau}, x^T\hat{\beta})d\left(\hat{F}_{0|\hat{\tau}}(x^T\hat{\beta}) - \hat{F}_0(x^T\hat{\beta}) - \left(F_{0|\hat{\tau}}(x^T\hat{\beta}) - F_0(x^T\hat{\beta})\right)\right)\right] \\
&+ \sqrt{n}\left[\int \hat{g}(\tau, x^T\hat{\beta})d(F_{0|\hat{\tau}}(x^T\hat{\beta}) - F_0(x^T\hat{\beta}))\right] \\
&= \Delta_{31} + \sqrt{n}\left[\int (\hat{g}(\hat{\tau}, x^T\hat{\beta}) - g(\tau, x^T\hat{\beta}))d(F_{0|\hat{\tau}}(x^T\hat{\beta}) - F_0(x^T\hat{\beta}))\right] \\
&+ \sqrt{n}\left[\int (g(\tau, x^T\hat{\beta}) - g(\tau, x^T\beta))d(F_{0|\hat{\tau}}(x^T\hat{\beta}) - F_0(x^T\hat{\beta}))\right] \\
&+ \sqrt{n}\left[\int g(\tau, x^T\beta)d(F_{0|\hat{\tau}}(x^T\hat{\beta}) - F_0(x^T\hat{\beta}))\right]
\end{aligned}$$

The first term will converge to zero by lemma 3. The second and third terms will converge to $o_p(1)$ by the proposition 2. For example,

$$\sqrt{n}\left[\int (g(\tau, x^T\hat{\beta}) - g(\tau, x^T\beta))d(F_{0|\hat{\tau}}(x^T\hat{\beta}) - F_0(x^T\hat{\beta}))\right] = O_p(n^{-1/2}h^{-1})O_p(1) = o_p(1)$$

By the asymptotic property of τ_0, the last term will be equal to

$$\sqrt{n}\left[\int g(\tau, x^T\beta)d(F_{0|\hat{\tau}}(x^T\hat{\beta}) - F_0(x^T\hat{\beta}))\right] = \sqrt{n}\left[\int g(\tau, x^T\beta)d(F'_{0,\tau_0}(x^T\beta, \tau_0)(\hat{\tau}_0 - \tau_0))/F(\tau_0)\right]$$

Next, we will move to Δ_1. In order to handle this term easily, we split the term into two parts:

$$
\begin{aligned}
\Delta_1 &= \sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - \hat{g}(\tau, x^T\hat{\beta})\right) d\hat{F}_0(x^T\hat{\beta})\right] \\
&= \underbrace{\sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - \hat{g}(\tau, x^T\hat{\beta})\right) d\left(\hat{F}_0(x^T\hat{\beta}) - F_0(x^T\beta)\right)\right]}_{\Delta_{11}} + \underbrace{\sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - \hat{g}(\tau, x^T\hat{\beta})\right) dF_0(x^T\beta)\right]}_{\Delta_{12}}
\end{aligned}
$$

Δ_{11} will vanish as n goes to infinity by proposition 2. The second term is a little bit complicated. By some algebraic operations, we have

$$
\begin{aligned}
\Delta_{12} =& \sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - \hat{g}(\tau, x^T\hat{\beta})\right) dF_0(x^T\beta)\right] \\
=& \sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - \hat{g}(\tau, x^T\hat{\beta})\right) - \left(g(\tau, x^T\hat{\beta}) - g(\tau, x^T\beta)\right) + \left(g(\tau, x^T\hat{\beta}) - g(\tau, x^T\beta)\right) dF_0(x^T\beta)\right] \\
=& \underbrace{\sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - g(\tau, x^T\hat{\beta})\right) dF_0(x^T\beta)\right]}_{\Delta_{12}^1} - \underbrace{\sqrt{n}\left[\int \left(\hat{g}(\tau, x^T\hat{\beta}) - g(\tau, x^T\beta)\right) dF_0(x^T\beta)\right]}_{\Delta_{12}^2} \\
& + \underbrace{\sqrt{n}\left[\int \left(g(\tau, x^T\hat{\beta}) - g(\tau, x^T\beta)\right) dF_0(x^T\beta)\right]}_{\Delta_{12}^3}
\end{aligned}
$$

The second term Δ_{12}^2 and Δ_{12}^{123} will be canceled out by the term T_{32}^2 below since $\Delta_{12}^{123} - \Delta_{12}^2 = -T_{32}^2$. Following the delta method and asymptotic property of β, Δ_{12}^3 will converge to a mean zero Gaussian process.

$$
\Delta_{12}^3 = \int g'_\beta(\tau, x^T\beta) dF_0(x^T\beta)\sqrt{n}(\hat{\beta} - \beta)
$$

We focus on analyzing the first term. For Δ_{12}^{1}, we can get as follows:

$$
\begin{aligned}
\Delta_{12}^{1} &= \sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - g(\hat{\tau}, x^T\hat{\beta})\right) + \left(g(\hat{\tau}, x^T\hat{\beta}) - g(\tau, x^T\hat{\beta})\right) dF_0(x^T\beta)\right] \\
&= \sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - g(\hat{\tau}, x^T\hat{\beta})\right) - \left(\hat{g}(\hat{\tau}, x^T\beta) - g(\hat{\tau}, x^T\beta)\right) dF_0(x^T\beta)\right] \\
&- \sqrt{n}\left[\int \left(\hat{g}(\tau, x^T\hat{\beta}) - g(\tau, x^T\hat{\beta}) - \left(\hat{g}(\tau, x^T\beta) - g(\tau, x^T\beta)\right)\right) dF_0(x^T\beta)\right] \\
&+ \underbrace{\sqrt{n}\left[\int \left(g(\hat{\tau}, x^T\hat{\beta}) - g(\tau, x^T\hat{\beta})\right) dF_0(x^T\beta)\right]}_{\Delta_{12}^{12}} \\
&+ \underbrace{\sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\beta) - g(\hat{\tau}, x^T\beta)\right) - \left(\hat{g}(\tau, x^T\beta) - g(\tau, x^T\beta)\right) + \left(\hat{g}(\tau, x^T\hat{\beta}) - g(\tau, x^T\hat{\beta})\right) dF_0(x^T\beta)\right]}_{\Delta_{12}^{13}} \\
&= A_1 - A_2 + \Delta_{12}^{12} + \Delta_{12}^{13}
\end{aligned}
$$

First, we will show that A_1 and A_2 will converge to $o_p(1)$ as n goes to infinity. For A_1, in order to simplify the analyzing process, we can rewrite as follows:

$$
\begin{aligned}
A_1 =& \sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - g(\hat{\tau}, x^T\hat{\beta})\right) d\left(F_0(x^T\beta) - F_0(x^T\hat{\beta})\right)\right] \\
&+ \sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\hat{\beta}) - g(\hat{\tau}, x^T\hat{\beta})\right) dF_0(x^T\hat{\beta}) - \int \left(\hat{g}(\hat{\tau}, x^T\beta) - g(\hat{\tau}, x^T\beta)\right) dF_0(x^T\beta)\right]
\end{aligned}
$$

The first term in A_1 will vanish as n goes to infinity by proposition 2 and Assumption 4. By lemma 4, the second term will converge to $o_p(1)$. The same process can be applied to A_2 which also converge to $o_p(1)$.

By the Taylor Expansion on Δ_{12}^{12} and Assumption 4, we have $g(\hat{\tau}, x^T\hat{\beta}) = g(\tau, x^T\beta) + g_\tau(\hat{\tau} - \tau) + g_\beta(\hat{\beta} - \beta)$ and $g(\tau, x^T\hat{\beta}) = g(\tau, x^T\beta) + g_\beta(\hat{\beta} - \beta)$. Then

$$
\Delta_{12}^{12} = \int g_\tau(\tau, x^T\beta) dF_0(x^T\beta)\sqrt{n}(\hat{\tau} - \tau)
$$

Move to Δ_{12}^{13}, it can be rewritten as follows:

$$\begin{aligned}
\Delta_{12}^{13} &= \sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\beta) - g(\hat{\tau}, x^T\beta)\right) - \left(\hat{g}(\tau, x^T\beta) - g(\tau, x^T\beta)\right) + \left(\hat{g}(\tau, x^T\hat{\beta}) - g(\tau, x^T\hat{\beta})\right) dF_0(x)\right] \\
&= \sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\beta) - g(\hat{\tau}, x^T\beta)\right) - \left(\hat{g}(\tau, x^T\beta) - g(\tau, x^T\beta)\right) dF_0(x)\right] \\
&+ \underbrace{\sqrt{n}\left[\int \left(\hat{g}(\tau, x^T\hat{\beta}) - g(\tau, x^T\hat{\beta})\right) dF_0(x)\right]}_{\Delta_{12}^{132}}
\end{aligned}$$

As we mentioned before, the second term Δ_{12}^{132} in the last equality will be canceled out. So we can only focus on the first term in the last equality above. Following the same notation in P.363 Bhattacharya and Mazumder (2011), we redefine some notations as follows: $z = x^T\beta$ and $\hat{z} = x^T\hat{\beta}$

$$\bar{m}(\tau, z) = \frac{1}{nh}\sum_i K(\frac{z_i - z}{h})1\{Y_i^{*c} \leq \tau_1, Y_i^{*p} \leq \tau_0\}$$

whose expectation is given by

$$\begin{aligned}
\bar{m}^*(\tau, z) &= E_{X_i}\left(\frac{1}{h}K(\frac{z_i - z}{h})\phi(Z_i, \tau_1, \tau_0)\right) \\
&= \int K(u)f(z + uh)\phi(z + uh, \tau_1, \tau_0)du \\
&= f(z)\phi(z, \tau_1, \tau_0) + O(h^r)
\end{aligned}$$

So we can get

$$\begin{aligned}
\bar{m}^*(\hat{\tau}, z) &= f(z)\phi(z, \hat{\tau}_1, \hat{\tau}_0) + O(h^r) \\
&= f(z)[\phi(z, \tau_1, \tau_0) + \phi_1(z, \tilde{\tau}_1, \tilde{\tau}_0)(\hat{\tau}_1 - \tau_1) + \phi_0(z, \tilde{\tau}_1, \tilde{\tau}_0)(\hat{\tau}_0 - \tau_0)] + O(h^r)
\end{aligned}$$

where $\tilde{\tau}$ denote value intermediate between $\hat{\tau}$ and τ. ϕ_1 and ϕ_0 are partial derivatives with respect to τ_1 and τ_0. ϕ is the conditional distribution of $1\{Y_i^{*c} \leq \tau_1, Y_i^{*p} \leq \tau_0\}$ on z. e.g

$Pr\left(Y_i^{*c} \le \tau_1, Y_i^{*p} \le \tau_0 \middle| z\right)$.

$$
\begin{aligned}
&\sqrt{n}\left[\int \left(\hat{g}(\hat{\tau}, x^T\beta) - g(\hat{\tau}, x^T\beta)\right) - \left(\hat{g}(\tau, x^T\beta) - g(\tau, x^T\beta)\right) dF_0(z)\right] \\
&= \sqrt{n}\int \frac{\frac{1}{nh}\sum_i 1\{Y_i^{*c} \le \hat{\tau}_1, Y_i^{*p} \le \hat{\tau}_0\}K(\frac{z_i - z}{h})}{\frac{1}{nh}\sum_i K(\frac{z_i-z}{h})1\{Y_i^{*p} \le \hat{\tau}_0\}} - g(\hat{\tau}, x^T\beta)dF_0(z) \\
&- \sqrt{n}\int \left(\frac{\frac{1}{nh}\sum_i 1\{Y_i^{*c} \le \tau_1, Y_i^{*p} \le \tau_0\}K(\frac{z_i - z}{h})}{\frac{1}{nh}\sum_i K(\frac{z_i-z}{h})1\{Y_i^{*p} \le \tau_0\}} - g(\tau, x^T\beta)\right) dF_0(z) \\
&= \sqrt{n}\int \left(\frac{\bar{m}(\hat{\tau}, z)}{\hat{f}(z, \hat{\tau}_0)} - \frac{f(z)\phi(z, \hat{\tau}_1, \hat{\tau}_0)}{f(z, \hat{\tau}_0)}\right) dF_0(z) - \sqrt{n}\int \left(\frac{\bar{m}(\tau, z)}{\hat{f}(z, \tau_0)} - \frac{f(z)\phi(z, \tau_1, \tau_0)}{f(z, \tau_0)}\right) dF_0(z)
\end{aligned}
$$

By lemma 3 and Assumption 7, we have that $O_p\left((nh)^{-1}\right) = o_p(n^{-1/2})$, then $\Delta_{12}^{13} = o_p(1)$.

Next, we consider the term Δ_2. By algebraic operations, we obtain

$$
\begin{aligned}
\Delta_2 &= \sqrt{n}\left[\int \hat{g}(x^T\hat{\beta})d\hat{F}_0(x^T\hat{\beta}) - g(x^T\beta)dF_0(x^T\beta)\right] \\
&= \sqrt{n}\left[\int \hat{g}(x^T\hat{\beta})d\hat{F}_0(x^T\hat{\beta}) - \int \hat{g}(x^T\beta)d\hat{F}_0(x^T\beta)\right] + \underbrace{\sqrt{n}\left[\int \hat{g}(x^T\beta)d\hat{F}_0(x^T\beta) - \int g(x^T\beta)dF_0(x^T\beta)\right]}_{T_1} \\
&= \underbrace{\sqrt{n}\left[\int \left(\hat{g}(x^T\hat{\beta}) - \hat{g}(x^T\beta)\right)\left(d\hat{F}_0(x^T\hat{\beta}) - \hat{F}_0(x^T\beta)\right)\right]}_{T_4} + \underbrace{\sqrt{n}\left[\int \left(\hat{g}(x^T\hat{\beta}) - \hat{g}(x^T\beta)\right) d\hat{F}_0(x^T\beta)\right]}_{T_3} \\
&+ \underbrace{\sqrt{n}\left[\int \hat{g}(x^T\beta)\left(d\hat{F}_0(x^T\hat{\beta}) - \hat{F}_0(x^T\beta)\right)\right]}_{T_2} + T_1 \\
&= T_4 + T_3 + T_2 + T_1
\end{aligned}
$$

By splitting into four parts, it is more easier for us to handle separately. Proposition 2 establishes that $T_4 = o_p(1)$ vanishes as n goes to infinity. Starting with T_2, we can obtain

$$
\begin{aligned}
T_2 =& \sqrt{n}\left[\int \hat{g}(x^T\beta)\left(d\hat{F}_0(x^T\hat{\beta}) - d\hat{F}_0(x^T\beta)\right)\right] \\
=& \sqrt{n}\left[\int g(x^T\beta)\left(d\hat{F}_0(x^T\hat{\beta}) - d\hat{F}_0(x^T\beta)\right)\right] + \sqrt{n}\left[\int \left(\hat{g}(x^T\beta) - g(x^T\beta)\right)\left(d\hat{F}_0(x^T\hat{\beta}) - d\hat{F}_0(x^T\beta)\right)\right] \\
=& T_{21} + T_{22}
\end{aligned}
$$

The second term can be divided into

$$
\begin{aligned}
T_{22} = \sqrt{n}\left[\int \left(\hat{g}(x^T\beta) - g(x^T\beta)\right)\left(d\hat{F}_0(x^T\hat{\beta}) - dF_0(x^T\hat{\beta}) - \left(d\hat{F}_0(x^T\beta) - dF_0(x^T\beta)\right)\right)\right] \\
+ \sqrt{n}\left[\int \left(\hat{g}(x^T\beta) - g(x^T\beta)\right)\left(dF_0(x^T\hat{\beta}) - dF_0(x^T\beta)\right)\right]
\end{aligned}
$$

We can show that this term converges to zero as n goes to infinity. In the first line, we can show the equation is equal to $o_p(1)$ by proposition 2. For the T_{21}, we can do this by using Taylor expansion as follows:

$$
\begin{aligned}
T_{21} = \sqrt{n}\left[\int g(x^T\beta)\left(d\hat{F}_0(x^T\hat{\beta}) - dF_0(x^T\hat{\beta}) - \left(d\hat{F}_0(x^T\beta) - dF_0(x^T\beta)\right)\right)\right] \\
+ \sqrt{n}\left[\int g(x^T\beta)\left(dF_0(x^T\hat{\beta}) - dF_0(x^T\beta)\right)\right]
\end{aligned}
$$

The first term will be $o_p(1)$ by lemma 3. The second term will be

$$
\sqrt{n}\left[\int g(x^T\beta)\left(dF'_{0\tau}(x^T\beta)(\hat{\beta} - \beta)/F(\tau)\right)\right]
$$

For the term T_3, we can show the asymptotic process as follows:

$$
\begin{aligned}
T_3 &= \sqrt{n}\left[\int \left(\hat{g}(x^T\hat{\beta}) - \hat{g}(x^T\beta)\right) d\hat{F}_0(x^T\beta)\right] \\
&= \sqrt{n}\left[\int \left(\hat{g}(x^T\hat{\beta}) - \hat{g}(x^T\beta)\right) d\left(\hat{F}_0(x^T\beta) - F_0(x^T\beta)\right)\right] + \sqrt{n}\left[\int \left(\hat{g}(x^T\hat{\beta}) - \hat{g}(x^T\beta)\right) dF_0(x^T\beta)\right] \\
&= T_{31} + T_{32}
\end{aligned}
$$

we can analyze T_3 separately. For the first term, we can rewrite it into

$$
\begin{aligned}
T_{31} &= \sqrt{n}\left[\int \left(\hat{g}(x^T\hat{\beta}) - \hat{g}(x^T\beta)\right) d\left(\hat{F}_0(x^T\beta) - F_0(x^T\beta)\right)\right] \\
&= \sqrt{n}\left[\int \left(\hat{g}(x^T\hat{\beta}) - g(x^T\hat{\beta}) - \left(\hat{g}(x^T\beta) - g(x^T\beta)\right)\right) d\left(\hat{F}_0(x^T\beta) - F_0(x^T\beta)\right)\right] \\
&+ \sqrt{n}\left[\int \left(g(x^T\hat{\beta}) - g(x^T\beta)\right) d\left(\hat{F}_0(x^T\beta) - F_0(x^T\beta)\right)\right]
\end{aligned}
$$

This term $\hat{g}(x^T\hat{\beta}) - g(x^T\hat{\beta}) - \left(\hat{g}(x^T\beta) - g(x^T\beta)\right)$ can be shown to converge to $O_p(n^{-\frac{1}{2}}h^{-1})$ for each x by lemma 3. And $\hat{F}_0(x^T\beta) - F_0(x^T\beta) = O_p(n^{-1/2})$ by the asymptotic theory of empirical process. Then we have $O_p(n^{-\frac{1}{2}}h^{-1})O_p(1) = o_p(1)$. The second term can be handled easily by taking the Taylor expansion inside the integral e.g $g^{'}(x^T\beta)(\hat{\beta}-\beta) = O_p\left((\sqrt{n})^{-1}\right) = o_p(1)$ and $\sqrt{n}\left(\hat{F}_0(x^T\beta) - F_0(x^T\beta)\right) = O_p(1)$.

Let us move to the next term.

$$
\begin{aligned}
T_{32} &= \sqrt{n}\left[\int \left(\hat{g}(x^T\hat{\beta}) - g(x^T\hat{\beta}) - \left(\hat{g}(x^T\beta) - g(x^T\beta)\right)\right) dF_0(x^T\beta)\right] + \sqrt{n}\left[\int \left(g(x^T\hat{\beta}) - g(x^T\beta)\right) dF_0(x^T\beta)\right] \\
&= T_{32}^1 + T_{32}^2
\end{aligned}
$$

where T_{32} will converge in distribution to a mean zero tight Gaussian process. The second term can be rewritten as follows:

$$
\begin{aligned}
T_{32}^2 &= \sqrt{n}\left[\int \left(g(x^T\hat{\beta}) - g(x^T\beta)\right) dF_0(x^T\beta)\right] \\
&= \left[\int g_\beta(x^T\beta) dF_0(x^T\beta)\right] \sqrt{n}\left(\hat{\beta}-\beta\right)
\end{aligned}
$$

where $g^1(x^T\beta)$ is the first order derivative with respect to β. The first term T_{32}^1 can be analyzed by using lemma 3

$$
T_{32}^1 = \sqrt{n}\int \left[\hat{g}(x^T\hat{\beta}) - \hat{g}(x^T\beta) - \left(g(x^T\hat{\beta}) - g(x^T\beta)\right)\right] dF_0(x^T\beta)
$$

Following lemma 3, we have that $T_{32}^1 = o_p(1)$ since the term T_{32}^1 satisfies the stochastic equicontinuity condition.

Finally, for the T_1 term, we can treat $\hat{g}$ as univariate nonparametric-kernel regression estimator, since we see β as known, we can treat single index as a single variable. For simplicity, we can replace $x^T\beta$ with z. We will show that this term will converge to a mean zero Gaussian process. Set $G_n g = \sqrt{n}(P_n - P)g$, $P_n g = \frac{1}{n}\sum g(z_i)$ and $Pg = \int g dP$.

$$
\begin{aligned}
T_1 =& \sqrt{n}\left[\int \hat{g}(z) d\hat{F}_0(z) - \int g(z) dF_0(z)\right] \\
=& G_n g + G_n(\hat{g} - g) + \sqrt{n}P(\hat{g} - g)
\end{aligned}
$$

By lemma 1, we can show the second term will be $o_p(1)$. The first term can be write as

$\frac{1}{\sqrt{n}} \sum_i (g(z_i) - E(g(z)))$, according to the assumptions we have made on the g($\cdot$) and Theorem 2.7.1 in VW, this term will converge to a mean zero Gaussian process. The last term will be

$$\frac{1}{\sqrt{n}} \sum_i^n (1\{Y_i^{*c}\} - g(z_i)) V(z_i, \tau_0) + O_p(h^r)$$

According to Assumption 7, $O_p(h^r) = o_p(n^{-1/2})$.

Lemma 1. *Under Assumptions 3 to 7, it holds that*

$$\sup_{z \in \mathcal{Z}} |G_n(\hat{g}(z) - g(z))| = o_p(1)$$

Proof. Let $h(z,g) = g(z)$ and $h(z,\hat{g}) = \hat{g}(z)$. Here we can treat z as a univariate variable. Then in order to prove the results above, we have to show that $\{h(\tilde{g}, z) : \tilde{g} \in H_0, z \in \mathcal{Z}\}$ is a $\mathcal{P}$-Donsker class and $\sup_z (h(z,\hat{g}) - h(z,g))^2 \xrightarrow{p} 0$. By Assumption 7 and proposition 2, we can show that $\sup_z (h(z,\hat{g}) - h(z,g))^2 = o_p(n^{-1/2})$ as n goes to infinity. By Assumptions 4 and 6, we know that $\{h(\tilde{g}, z) : \tilde{g} \in H_0, z \in \mathcal{Z}\}$ is a class of continuous functions defined on compact sets with uniform bounded derivatives and thus satisfy the definition of smooth functions in 2.7.1 in Li and Racine (2007). And it is a $\mathcal{P}$-Donsker class. By Theorem 2.1 in Vaart, Wellner, et al. (2007), the results follows. □

Lemma 2. *Let $V_n(\theta)$ be a stochastic process indexed by $\theta \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is a compact subset of $\mathbb{R}^q$. If for all θ_1 and θ_2, we have*

$$E\left[|V_n(\theta_1) - V_n(\theta_2)|^\alpha\right] \leq C\,|\theta_1 - \theta_2|^\gamma$$

for some $\alpha > 0$,$\gamma > 1$ and $C > 0$ is a constant, then $V_n(\theta)$ is stochastically equicontinuous for $\theta \in \mathbf{\Theta}$.

Proof. It can be shown by referring to the Theorem A.8 in Li and Racine (2007). □

Lemma 3. *Under Assumptions 3 to 7 and proposition 2*

$$(1)\sup|\hat{g}(\hat{z})-g(\hat{z})-(\hat{g}(z)-g(z))|=O_p\left((n)^{-1}h^{-2}\right)$$
$$(2)\sup|\hat{F}_{0|\hat{\tau}}(x^T\hat{\beta})-\hat{F}_0(x^T\hat{\beta})-\left(F_{0|\hat{\tau}}(x^T\hat{\beta})-F_0(x^T\hat{\beta})\right)|=o_p(n^{-1/2})$$
$$(3)\sqrt{n}\int\left[\hat{g}(x^T\hat{\beta})-\hat{g}(x^T\beta)-\left(g(x^T\hat{\beta})-g(x^T\beta)\right)\right]dF_0(x^T\beta)=o_p(1)$$

Proof. We set $\hat{z}=x^T\hat{\beta}$, $z=x^T\beta$, $\bar{m}(\hat{z})=\hat{g}(\hat{z})$, $\bar{m}^*(\hat{z})=E\left(\hat{g}(\hat{z})\right)$, $\bar{m}(z)=\hat{g}(z)$ and $\bar{m}^*(z)=E\left(\hat{g}(z)\right)$.

$$\bar{m}(z)=\frac{1}{nh}\sum_i K(\frac{z_i-z}{h})1\{Y_i^{*c}\le\tau_1,Y_i^{*p}\le\tau_0\}$$

whose expectation is given by

$$\bar{m}^*(z)=E_{Z_i}\left(\frac{1}{h}K(\frac{z_i-z}{h})\phi(z_i)\right)=\int K(u)f(z+uh)\phi(z+uh)du=f(z)\phi(z)+O(h^r)$$

So we can get

$$\bar{m}^*(\hat{z})=f(\hat{z})\phi(\hat{z})+O(h^r)=f(z)\phi(z)+\left(f(z)\phi'(z)+f'(z)\phi(z)\right)(\hat{z}-z)+O(h^r)$$

Following the notation we set before, We can rewrite the equation as follows

$$\begin{aligned}\hat{g}(\hat{z})-g(\hat{z})-(\hat{g}(z)-g(z))=&\frac{\bar{m}(\hat{z})-\bar{m}^*(\hat{z})-(\bar{m}(z)-\bar{m}^*(z))}{f(z,\tau_0)}+\frac{\bar{m}(\hat{z})}{(f(\hat{z},\tau_0))^2}(f(\hat{z},\tau_0)-\hat{f}(\hat{z},\tau_0))\\&+\frac{\bar{m}(z)}{(f(z),\tau_0))^2}(f(z,\tau_0)-\hat{f}(z,\tau_0))+\left(\frac{\bar{m}(\hat{z})-\bar{m}^*(\hat{z})}{(f(z,\tau_0))^2}\right)(f(z,\tau_0)-f(\hat{z},\tau_0))\\&+O_p\left(n^{-1}h^{-2}\right)\end{aligned}$$

We use geometric expansion[7] to the denominator and undersmooth kernel function to eliminate the bias term. By the uniform boundedness of $\frac{1}{h}K(\cdot)$ and $g(\cdot)$, plus the compactness of set of the independent variables, we see the first term will converge to $o_p(1)$ since this term meets the definition of classes of functions, which satisfy the stochastic equicontinuity in Andrews (1994). The rest will follow the same process by common kernel smoothing theory proposition 2 and

[7]For example, we want to expand the $1/\hat{f}$, then by using geometric expansion, $\frac{1}{\hat{f}}=\frac{1}{f}(1-\Delta)^{-1}=\frac{1}{f}(1+\Delta+\Delta^2(1-\Delta)^{-1})$, where $\Delta=\frac{f-\hat{f}}{f}$.

Assumptions 4, 6 and 7.

$$
\begin{aligned}
&\sup|\frac{\bar{m}(z)}{(f(z),\tau_0))^2}(f(z,\tau_0)-\hat{f}(z,\tau_0))| \\
&=\sup|\frac{\frac{1}{nh}\sum_i K(\frac{z_i-z}{h})1\{Y_i^{*c}\le\tau_1,Y_i^{*p}\le\tau_0\}}{(f(z),\tau_0))^2}(f(z,\tau_0)-\hat{f}(z,\tau_0))| \\
&\le\sup\left|c\frac{\frac{1}{\sqrt{n}}\sum_i 1\{Y_i^{*c}\le\tau_1,Y_i^{*p}\le\tau_0\}}{(f(z),\tau_0))^2}\right|\sup\left|(n^{-1/2}h^{-1})(f(z,\tau_0)-\hat{f}(z,\tau_0))\right| \\
&=O_p(1)O_p(n^{-1}h^{-2})=o_p(n^{-1/2})
\end{aligned}
$$

Where $|K(\cdot)|<c$ is uniformly bounded. The last equality can be obtained by proposition 2, Assumption 6 and uniform convergence rate of indicator function which can be obtained by using Hoeffding's inequality such as $P(\sup|P_n(A)-P(A)|>t)\le cexp(-2nt^2)$.

Next, we prove the third term. Inserting $\bar{m}(z)$, $\bar{m}(\hat{z})$, $\bar{m}^*(z)$ and $\bar{m}^*(\hat{z})$ into the function below and splitting the third term in the lemma into 3 parts, we can get

$$
\begin{aligned}
&\sqrt{n}\int\left[\hat{g}(x^T\hat{\beta})-\hat{g}(x^T\beta)-\left(g(x^T\hat{\beta})-g(x^T\beta)\right)\right]dF_0(x^T\beta) \\
=&\sqrt{n}\int\left[\hat{g}(x^T\hat{\beta})-g(x^T\hat{\beta})\right]dF_0(x^T\hat{\beta})-\sqrt{n}\left[\int\left(\hat{g}(x^T\beta)-g(x^T\beta)\right)dF_0(x^T\beta)\right] \\
&+\sqrt{n}\left[\int\left(\hat{g}(x^T\hat{\beta})-g(x^T\hat{\beta})\right)d\left(F_0(x^T\beta)-F_0(x^T\hat{\beta})\right)\right]
\end{aligned}
$$

The first and term term converge to zero by lemma 4. The third term is $o_p(1)$ by proposition 2 and bandwidth specification in Assumption 7. Thus we can get the result that

$$
\sqrt{n}\int\left[\hat{g}(x^T\hat{\beta})-\hat{g}(x^T\beta)-\left(g(x^T\hat{\beta})-g(x^T\beta)\right)\right]dF_0(x^T\beta)=o_p(1)
$$

.

For (2), since the $\hat{f}(\hat{\tau}_0)=f(\tau_0)=O(1)$ is a constant number, we can just focus on the numerator without loss of generality. Define a class of functions as $\mathcal{F}=\{H_{\tau_0 z}=1_{(0,\tau_0]\times(0,z]}:\tau_0 z\in\mathcal{TZ}\}$. Then we can rewrite (2) as: $G_n(H(\hat{\tau}_0,\hat{z}))-H(\tau_0,z))$. According to Assumption 3, we have that $\mathcal{TZ}$ is compact since the Cartesian product of two compact sets is also compact. Then we can have a finite number of open balls to cover $\mathcal{TZ}$, which will leads to the situation that the minimum number of ϵ-bracket in $L_1(P)$ to cover $\mathcal{F}$ is finite — $N_{[]}(\epsilon,\mathcal{F},L_1(P))<\infty$. So $\mathcal{F}$ is Donsker class, we obtain that $\sup|G_n(H(\hat{\tau}_0,\hat{z}))-H(\tau_0,z))|=o_p(1)$ by stochastic equicontinuity which means (3) holds.

$\square$

Lemma 4. *Under Assumptions 3 to 7, we have that*

$$\sqrt{n}P_{\hat{Z}}\left(\hat{g}(\hat{z}) - g(\hat{z})\right) - \sqrt{n}P_Z\left(\hat{g}(z) - g(z)\right) = o_p(1)$$

where $\sqrt{n}P_{\hat{Z}}\left(\hat{g}(\hat{z}) - g(\hat{z})\right) = \sqrt{n}\int\left(\hat{g}(\hat{z}) - g(\hat{z})\right)dF_0(\hat{z})$ *and* $\sqrt{n}P_Z\left(\hat{g}(z) - g(z)\right) = \sqrt{n}\int\left(\hat{g}(z) - g(z)\right)dF_0(z)$

Proof. This Lemma can be proved by using the stochastic equicontinuity. Lemma 3 provide a sufficient condition for stochastic equicontinuity. Just for simplicity, we only focus on $\sqrt{n}P_Z\left(\hat{g}(z) - g(z)\right)$, the other one will follow the same process.

In order to simplify the proof, We drop $\sqrt{n}$ for simplicity and split the term into two parts:

$$\begin{aligned}
P_Z\left(\hat{g}(z) - g(z)\right) &= \int \frac{1}{n}\left(\frac{\frac{1}{h}\sum_i 1\{Y_i^{*c} \le \tau_1, Y_i^{*p} \le \tau_0\}K(\frac{z_i - z}{h})}{\hat{f}(z,\tau_0)} - g(z)\right)dF_0(z) \\
&= \int \frac{1}{n}\left(\frac{\frac{1}{h}\sum_i 1\{Y_i^{*c} \le \tau_1\}1\{Y_i^{*p} \le \tau_0\}K(\frac{z_i - z}{h})}{\frac{1}{nh}\sum_i K(\frac{z_i - z}{h})1\{Y_i^{*p} \le \tau_0\}} - \frac{g(z)\hat{f}(z,\tau_0)}{\hat{f}(z,\tau_0)}\right)dF_0(z) \\
&= \int \frac{1}{n}\left(\frac{\frac{1}{h}\sum_i \left(1\{Y_i^{*c} \le \tau_1\} - g(z)\right)1\{Y_i^{*p} \le \tau_0\}K(\frac{z_i - z}{h})}{\hat{f}(z,\tau_0)}\right)\frac{f^c(z,\tau_0)}{f^c(\tau_0)}dz \\
&= \frac{1}{n}\int\left(\frac{\frac{1}{h}\sum_i \left(1\{Y_i^{*c} \le \tau_1\} - g(z)\right)1\{Y_i^{*p} \le \tau_0\}K(\frac{z_i - z}{h})}{f(z,\tau_0)}\right)\frac{f^c(z,\tau_0)}{f^c(\tau_0)}dz \\
&- \frac{1}{n}\int\left(\frac{\frac{1}{h}\sum_i \left(1\{Y_i^{*c} \le \tau_1\} - g(z)\right)1\{Y_i^{*p} \le \tau_0\}K(\frac{z_i - z}{h})}{f(z,\tau_0)}\right)\frac{f^c(z,\tau_0)\left(\hat{f}(z,\tau_0) - f(z,\tau_0)\right)}{f(z,\tau_0)f^c(\tau_0)}dz \\
&+ o_p(n^{-\frac{1}{2}}) \\
&= \mathbf{H}_1 + \mathbf{H}_2 + o_p(n^{-\frac{1}{2}})
\end{aligned}$$

We get $o_p(n^{-\frac{1}{2}})$ since $\hat{f} - f = o_p(n^{-\frac{1}{4}})$ uniformly on $\mathcal{ZT}$. We first consider the first term. By using standard kernel smoothing techniques, we have

$$\begin{aligned}
\mathbf{H}_1 &= \frac{1}{n}\int \left(\frac{1}{h}\sum_i \left(1\{Y_i^{*c} \le \tau_1\}1\{Y_i^{*p} \le \tau_0\} - g(z)\right) K(\frac{z_i - z}{h})\right) dz \\
&= \frac{1}{n}\int \left(\frac{1}{h}\sum_i (1\{Y_i^{*c} \le \tau_1\} - g(z)) K(\frac{z_i - z}{h})\right) V(z,\tau_0) dz \\
&= \frac{1}{n}\int \left(\sum_i (1\{Y_i^{*c} \le \tau_1\} - g(z_i + h\mu)) K(\mu)\right) V(z_i + h\mu, \tau_0) d\mu \\
&= \frac{1}{n}\sum_i (1\{Y_i^{*c}\} - g(z_i)) V(z_i,\tau_0) + O_p(h^r)
\end{aligned}$$

where $V(z,\tau_0) = \frac{1\{Y_i^{*p} \le \tau_0\} f^c(z,\tau_0)}{f(z,\tau_0) f^c(\tau_0)}$ and $F_0(z) = \int f^c(z,\tau_0)/f^c(\tau_0) dz$ is the distribution of another group of people. The third and last equality follows by using change of variables, interchanging the order of integration and summation and using the kernel properties. By Assumption 7, $O_p(h^r) = o_p(n^{-\frac{1}{2}})$.

For the second term $\mathbf{H}_2$, by the same process we have done on $\mathbf{H}_1$, we have

$$\begin{aligned}
&\frac{1}{n}\int \left(\frac{1}{h}\sum_i (1\{Y_i^{*c} \le \tau_1\} - g(z)) 1\{Y_i^{*p} \le \tau_0\} K(\frac{z_i - z}{h})\right) \frac{f^c(z,\tau_0)\left(\hat{f}(z,\tau_0) - f(z,\tau_0)\right)}{f(z,\tau_0)^2 f^c(\tau_0)} dz \\
&= \frac{1}{n}\int \left(\frac{1}{h}\sum_i (1\{Y_i^{*c} \le \tau_1\} - g(z)) K(\frac{z_i - z}{h})\right) v(z,\tau)\left(\hat{f}(z,\tau_0) - f(z,\tau_0)\right) dz \\
&= \frac{1}{n^2}\sum_{i,j} v(z_i,\tau)(1\{Y_i^{*c} \le \tau_1\} - g(z_i))\left(1\{Y_i^{*p} \le \tau_0\}\frac{1}{h}K(\frac{z_j - z_i}{h}) - f(z_i,\tau_0)\right) + O_p(h^r) \\
&= U_n(\tau) + O_p(h^2)
\end{aligned}$$

where $v(z,\tau) = 1\{Y_i^{*p} \le \tau_0\} f^c(z,\tau_0)/f(z,\tau_0)^2 f^c(\tau_0)$. Let

$$U_n(\tau) = v(z_i,\tau)(1\{Y_i^{*c} \le \tau_1\} - g(z_i))\left(1\{Y_i^{*p} \le \tau_0\}\frac{1}{h}K(\frac{z_j - z_i}{h}) - \bar{f}(z_i,\tau_0)\right) + o_p(n^{-1/2})$$

Where $\bar{f}(z,\tau_0) = E_z\left(1\{Y_i^{*p} \le \tau_0\}\frac{1}{h}K(\frac{z_i - z}{h})\right)$ and $\sup|\bar{f} - f| = O_p(h^2) = o_p(n^{-1/2})$. As stated in Rothe (2010), we can further write as

$$U_n(\tau) = \frac{1}{n^2}\sum_i\sum_{j\neq i} H(Z_i, Z_j; \tau, h) + \frac{1}{n^2}\sum_i H(Z_i, Z_i; \tau, h) + o_p(n^{-1/2})$$

The first term satisfy the definition of a second order degenerate U-statistics since $E(H|Z_i) = 0$

and H is symmetric. By Lemma 2 in Rothe (2010), we can show that the class of functions $\mathcal{H} = \{hH(\cdot;\tau,h)\}$ is uniformly bounded and has a constant envelop Corollary 4 in Sherman (1994), we can conclude that

$$\sup_{\tau\in\mathcal{T}} \left| \frac{1}{n^2} \sum_i \sum_{j\neq i} H(Z_i, Z_j; \tau, h) \right| = O_p\left((nh)^{-1}\right) = o_p(n^{-1/2})$$

And the second term also converge to $O_p\left((nh)^{-1}\right) = o_p(n^{-1/2})$, thus $\mathbf{H}_2 = o_p(n^{-1/2})$ by Assumption 7.

By the same process we can get

$$N_n(\hat{z},\tau) = \sqrt{n} P_{\hat{Z}}\left(\hat{g}(\hat{z}) - g(\hat{z})\right) = \frac{1}{\sqrt{n}} \sum_i \left(1\{Y_i^{*c}\} - g(\hat{z}_i)\right) V(\hat{z}_i, \tau_0) + o_p(1)$$
$$N_n(z,\tau) = \sqrt{n} P_{Z}\left(\hat{g}(z) - g(z)\right) = \frac{1}{\sqrt{n}} \sum_i \left(1\{Y_i^{*c}\} - g(z_i)\right) V(z_i, \tau_0) + o_p(1)$$

By uniform boundedness of f g and their derivatives, lemma 2 shows that $N_n(\cdot,\tau)$ is stochastically equicontinuous for all $z\tau \in \mathcal{Z}\mathcal{T}$ when $\alpha = 1$ and $\gamma = 2$. Thus we can conclude that

$$\sup_{z\tau\in\mathcal{Z}\mathcal{T}} |N_n(\hat{z},\tau) - N_n(z,\tau)| = o_p(1)$$

□

Proposition 2. *Under Assumptions 3 to 7, we have that*

1. $\sup_{\tilde{z}\in\mathcal{Z}} |\hat{g}_{\tau,z}(\tilde{\tau},\tilde{z}) - g_{\tau,z}(\tilde{\tau},\tilde{z})| = O_p(h^{-1}n^{-1/2})$
2. $\sup_{\tilde{z}\in\mathcal{Z}} |\hat{g}(\hat{\tau},\tilde{z}) - g(\tau,\tilde{z})| = O_p(h^{-1}n^{-1/2})$
3. $\sup_{\tilde{z}\in\mathcal{Z}} |\hat{g}(\hat{\tau},\tilde{z}) - \hat{g}(\tau,\tilde{z})| = O_p(n^{-1/2})$
4. $\sup_{\tilde{z}\in\mathcal{Z},\tilde{\tau}\in\mathcal{T}} |\hat{g}(\tilde{\tau},\tilde{z}) - g(\tilde{\tau},\tilde{z})| = O_p(h^{-1}n^{-1/2})$
5. $\sup_{x\in\mathcal{X},\|\hat{\beta}-\beta\|\leq\epsilon_n} |\hat{g}(\hat{z}) - \hat{g}(z)| = O_p(n^{-1/2})$
6. $\sup_{\tilde{z}\in\mathcal{Z},\|\hat{\beta}-\beta\|\leq\epsilon_n} |\hat{f}(\tilde{z},\tilde{\tau}) - f(\tilde{z},\tilde{\tau})| = O_p(h^{-1}n^{-1/2})$
7. $\sup_{\hat{z}\in\mathcal{Z},\|\hat{\beta}-\beta\|\leq\epsilon_n} |\hat{F}_0(\hat{z}) - \hat{F}_0(z)| = O_p(1/\sqrt{n})$
8. $\sup_{x\in\mathcal{X},\|\hat{\beta}-\beta\|\leq\epsilon_n} |\hat{f}(\tilde{z},\tilde{\tau}) - f(z,\tilde{\tau})| = O_p(h^{-1}n^{-1/2})$

where $\epsilon_n = o(1)$ as n goes to infinity.

Proof. We consider the 1 and 6 first. $g_{\tau,z} = g(\tau, z)f(z, \tau)$ corresponds to the true value of the numerator of kernel-based estimated conditional distribution, we show that $|\hat{g}_{\tau,z} - g_{\tau,z}| = O_p(n^{-1/2}h^{-1})$ is uniformly on τ, β and z. By the bias-reducing assumption, we can only focus on $|\hat{g}_{\tau,z} - E(\hat{g}_{\tau,z})|$. $\mathcal{Z}\mathcal{T}$ is a compact Cartesian product since $\mathcal{Z}$ and $\mathcal{T}$ are compact sets, thus we can cover it by a class of finite number of open balls—$\mathcal{C}_N = \{C_{jN} : j = 1, \cdots, b_N\}$. Each ball has the same radius r_N and centered at w_{jN}. For $w_{jN} \in C_{jN}$, we obtain

$$\begin{aligned}
\sup_{w \in \mathcal{Z}\mathcal{T}} |\hat{g}_{\tau,z}(w) - E(\hat{g}_{\tau,z}(w))| \le \max_j \sup_{w \in C_{jN}} & [|\hat{g}_{\tau,z}(w) - \hat{g}_{\tau,z}(w_{jN})| \\
& + |\hat{g}_{\tau,z}(w_{jN}) - E(\hat{g}_{\tau,z}(w_{jN}))| \\
& + |E(\hat{g}_{\tau,z}(w_{jN})) - E(\hat{g}_{\tau,z}(w))|] \\
= B_1 + B_2 + B_3 &
\end{aligned}$$

By Assumption 6 and Mean Value Theorem, we obtain that $\hat{g}(z)$ satisfy Lipschitz condition since $|\hat{g}(z_1) - \hat{g}(z_2)| \le c\|z_1 - z_2\|h^{-1}$ where $\|h\partial_z \hat{g}(\tilde{z})\| < c$, then $\max_j \sup_{w \in C_{jN}} h^{-1} |\hat{g}_{\tau,z}(w) - \hat{g}_{\tau,z}(w_{jN})| = c\, r_N$ since $\max_j \sup_{w \in C_{jN}} \|w - w_{jN}\| = r_N$. The third term follows from B_1 and will also be equal to $c\, r_N/h$. For the term B_2, we have

$$\begin{aligned}
P(\max_j |\hat{g}_{\tau,z}(w_{jN}) - E(\hat{g}_{\tau,z}(w_{jN}))| > \delta_n) &\le P\left(\bigcup_j^{b_N} (|\hat{g}_{\tau,z}(w_{jN}) - E(\hat{g}_{\tau,z}(w_{jN}))| > \delta_n)\right) \\
&\le \sum_j^{b_N} P\left(h^{-1} |\hat{g}_{\tau,z}(w_{jN}) - E(\hat{g}_{\tau,z}(w_{jN}))| > h^{-1}\delta_n\right)
\end{aligned}$$

By using Hoeffding's inequality and Borel-Cantelli lemma, we can get

$$\begin{aligned}
P(\max_j |\hat{g}_{\tau,z}(w_{jN}) - E(\hat{g}_{\tau,z}(w_{jN}))| > \delta_n) &\le \sum_j^{b_N} P\left(h^{-1} |\hat{g}_{\tau,z}(w_{jN}) - E(\hat{g}_{\tau,z}(w_{jN}))| > h^{-1}\delta_n\right) \\
&\le 2b_N exp\left(-Dh^2 N\delta_N^2\right)
\end{aligned}$$

where $\delta_n = N^{-1/2}h^{-1}$ and D is a positive constant. Thus $B_2 = O_p(N^{-1/2}h^{-1})$. By setting

$r_N = \sqrt{N}$, we can get the other two terms will also be $O_p(N^{-1/2}h^{-1})$. By the same process, we can prove that $\sup_{\tilde{z}\in\mathcal{Z},\|\hat{\beta}-\beta\|\le\epsilon_n} |\hat{f}(\tilde{z},\tilde{\tau}) - \hat{f}(\tilde{z},\tilde{\tau})| = O_p(h^{-1}n^{-1/2})$.

Then from the proposition 1 and 6 which have been proved above, we can prove the fourth one. Since $\hat{g}(\hat{\tau},\tilde{z}) = \hat{g}_{\tau,z}(\hat{\tau},\tilde{z})/\hat{f}(\tilde{z},\hat{\tau}_0)$ and $g(\tau,\tilde{z}) = g_{\tau,z}(\hat{\tau},\tilde{z})/f(\tilde{z},\hat{\tau}_0)$, then we can have

$$
\begin{aligned}
\sup_{\tilde{z}\in\mathcal{Z}} |\hat{g}(\tilde{\tau},\tilde{z}) - g(\tau,\tilde{z})| &= \sup_{\tilde{z}\in\mathcal{Z}} \left| \frac{\hat{g}_{\tau,z}(\tilde{\tau},\tilde{z})}{\hat{f}(\tilde{z},\tilde{\tau}_0)} - \frac{g_{\tau,z}(\tilde{\tau},\tilde{z})}{f(\tilde{z},\tilde{\tau}_0)} \right| \\
&\le \sup_{\tilde{z}\in\mathcal{Z}} \left| \frac{\hat{g}_{\tau,z}(\tilde{\tau},\tilde{z}) - g_{\tau,z}(\tilde{\tau},\tilde{z})}{f(\tilde{z},\tilde{\tau}_0)} \right| + \sup_{\tilde{z}\in\mathcal{Z}} \left| \frac{\hat{g}_{\tau,z}(\tilde{\tau},\tilde{z}) \left(f(\tilde{z},\tilde{\tau}_0) - \hat{f}(\tilde{z},\tilde{\tau}_0) \right)}{f(\tilde{z},\tilde{\tau}_0)^2} \right| + O_p(n^{-1}h^{-2}) \\
&= O_p(n^{-1/2}h^{-1})
\end{aligned}
$$

For the third one, stochastic equicontinuity and Assumption 6 establish the third one. Here we will omit the symbol $\tilde{\cdot}$ for simplicity. Following the notation we have used above, we can get

$$
\begin{aligned}
\hat{g}(\hat{\tau},z) - \hat{g}(\tau,z) &= \frac{\frac{1}{nh}\sum_i 1\{Y_i^{*c} \le \hat{\tau}_1, Y_i^{*p} \le \hat{\tau}_0\} K(\frac{z_i - z}{h})}{\hat{f}(z,\hat{\tau}_0)} - \frac{\frac{1}{nh}\sum_i 1\{Y_i^{*c} \le \tau_1, Y_i^{*p} \le \tau_0\} K(\frac{z_i - z}{h})}{\hat{f}(z,\tau_0)} \\
&= \frac{V_{1N} - V_{2N}}{\hat{f}(z,\tau_0)} + \frac{V_{1N}(\hat{f}(z,\tau_0) - \hat{f}(z,\hat{\tau}_0))}{\hat{f}(z,\tau_0)^2} + \frac{\bar{m}^*(\hat{\tau},z) - \bar{m}^*(\tau,z)}{\hat{f}(z,\tau_0)} \\
&+ \frac{\bar{m}^*(\hat{\tau},z)(\hat{f}(z,\tau_0) - \hat{f}(z,\hat{\tau}_0))}{\hat{f}(z,\tau_0)} + O_p(1/n)
\end{aligned}
$$

Where $V_{1N} = \bar{m}(\hat{\tau},z) - \bar{m}^*(\hat{\tau},z)$ and $V_{2N} = \bar{m}(\tau,z) - \bar{m}^*(\tau,z)$. We can get $\sup|\hat{f}(z,\hat{\tau}_0) - \hat{f}(z,\tau_0)| = O_p(n^{-1/2})$ by setting $\bar{m}(\tau,z) = \frac{1}{nh}\sum_i K(\frac{z_i - z}{h}) 1\{Y_i^{*p} \le \tau_0\}$ and following the same process above. By Assumptions 3, 4 and 6, we know $\frac{1}{h}K(\cdot)1\{\cdot\}$ and the derivative $\phi_j(\cdot)$ are uniformly bounded. By definition of type IV function in Andrews (1994) and Bhattacharya and Mazumder (2011), $V_{1N} - V_{2N} = o_p(1)$ since it satisfies the entropy condition and thus is stochastic equicontinuity. The rest terms will all be $O_p(n^{-1/2})$ since $\sup|\hat{f}(z,\hat{\tau}_0) - \hat{f}(z,\tau_0)| = O_p(n^{-1/2})$ and $\hat{\tau} - \tau = O_p(n^{-1/2})$.

The second one will be followed by using triangle inequality, 2 and 4. Following the same proof process on 3, we can prove equation 5 by Assumptions 3, 4 and 6.

The equation 7 can be proved by using triangle inequality, stochastic equicontinuity of empirical process and uniform boundedness of derivative of F with respect to z.

$$\sup_{\hat{z}\in\mathcal{Z},\|\hat{\beta}-\beta\|\leq\epsilon_n} |\hat{F}_0(\hat{z}) - \hat{F}_0(z)| \leq \sup_{\hat{z}\in\mathcal{Z},\|\hat{\beta}-\beta\|\leq\epsilon_n} |\hat{F}_0(\hat{z}) - F_0(\hat{z}) - [\hat{F}_0(z) - F_0(z)]| + \sup_{\hat{z}\in\mathcal{Z},\|\hat{\beta}-\beta\|\leq\epsilon_n} |F_0(\hat{z}) - F_0(z)|$$
$$= O_p(1/\sqrt{n})$$

□

Proof of Theorem 2. We need to make some assumptions before proving the asymptotic theory.

Asymptotic theory for the coefficient β. Under some conditions mentioned in Chen, Linton, and Van Keilegom (2003), we will show that $\sqrt{n}(\hat{\beta} - \beta)$ will converge to a mean zero Gaussian process. Rothe (2009) uses the similar arguments to prove the semiparametric estimation of binary response models with endogenous variables. Although we both use maximum likelihood estimation which is based on nonparametric estimator, the coefficients in our paper are related to the quantile estimators (τ) which is the threshold of children's and parent's income, and thus are very different from the counterparts in Rothe (2009). By using MLE, we can obtain

$$\mathcal{L}_n(\beta) = \frac{1}{n}\sum_{i}^{n}\left[1\{Y_i^{*c} \leq \hat{\tau}_1\}\ln\left[\hat{G}(\hat{\tau}, X_i^T\beta)\right] + (1 - 1\{Y_i^{*c} \leq \hat{\tau}_1\})\ln\left[1 - \hat{G}(\hat{\tau}, X_i^T\beta)\right]\right]$$

If we choose a $\beta \in \mathcal{B}$ which will maximize $\mathcal{L}_n(\beta)$ which in turn will make $\partial_\beta\mathcal{L}_n(\beta) = 0$, then β is the Z-estimator. Define $M_n(\beta, \hat{h}) = \partial_\beta\mathcal{L}_n(\beta)$ where $\hat{h} = \hat{h}_{\beta,\hat{\tau}} = (\partial_\beta\hat{G}(\hat{\tau}, X_i^T\beta), \hat{G}(\hat{\tau}, X_i^T\beta), \hat{\tau}), \hat{h}_{\beta,\tau^0} = (\partial_\beta\hat{G}(\tau^0, X_i^T\hat{\beta}), \hat{G}(\tau^0, X_i^T\hat{\beta}), \tau^0)$ and $h_0 = h_{\beta_0,\tau^0} = (\partial_\beta G(\tau^0, X_i^T\beta_0), G(\tau^0, X_i^T\beta_0), \tau^0)$. Then for any h, we have

$$M_n(\beta, \hat{h}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\left[\partial_\beta\hat{G}(\hat{\tau}, X_i^T\beta)\right]\left(Y_i - \hat{G}(\hat{\tau}, X_i^T\beta)\right)}{\hat{G}(\hat{\tau}, X_i^T\beta)\left(1 - \hat{G}(\hat{\tau}, X_i^T\beta)\right)}$$

where $Y_i = 1\{Y_i^{*c} \leq \hat{\tau}_1\}$ and $M_n(\beta, \hat{h})$ is the score of likelihood function $\mathcal{L}_n(\beta)$. And, define the population moment function as $M(\beta, \hat{h}) = E(M_n(\beta, \hat{h}))$. Finally, following the definition in Chen, Linton, and Van Keilegom (2003) and the assumptions we made in this paper, we can define the space $\mathcal{H}$ for h which is a vector of functions with sup-norm $\| \|_\infty$. Set $\mathcal{H} = \mathcal{H}_1 \times \mathcal{T}$. By Assumptions 3, 4 and 6, $\mathcal{H}_1 = \{f, \partial^\alpha f : \mathbb{R} \mapsto \mathbb{R} : f, \partial^\alpha f \in C_C^r(\mathbb{Z}, \mathcal{T})\}$ where $C_C^r(\mathbb{X})$ is a class of continuous functions defined on $\mathbb{X}$ which are bounded by constant number C and have partial derivatives up to order r, and $\mathcal{T} = \{\tau : \tau = \sup_{y \in \mathcal{Y}}\{\hat{F}(y_i) \leq s : i = 1, 2, \cdots, n\}\}$ where $\hat{F}$ is the empirical distribution function and s is the sth quartile. By Dvoretzky-Kiefer-Wolfowith (DKW) inequality and Borel-Cantelli lemma, $\sup|\hat{F} - F| = O_p(n^{-1/3}) = o_p(n^{-1/4})$. Thus we can define $\| \, \|_{\mathcal{H}} = \| \, \|_\infty = \sup_{\beta \in \mathcal{B}} \max\{\|h_1\|_\infty, \|h_2\|_\infty, \|h_3\|_\infty\}$ where we use h_j to represent each element of vector h.

First, we are going to show that the conditions (2.1),(2.2) and (2.4) of Theorem 2 in Chen, Linton, and Van Keilegom (2003) can be satisfied and thus the asymptotic normality of the coefficients we are estimated can be proved. Since we have already mentioned before, $\hat{\beta}$ is the Z-estimator which make $\partial_\beta \mathcal{L}_n(\beta) = 0$, the condition $\|M_n(\hat{\beta}, \hat{h})\|_\infty \leq \inf_{\beta \in \mathcal{B}} \|M(\beta, \hat{h})\|_\infty + o_p(n^{-1/2})$ is satisfied. By Assumption 4, we can show that $\partial_\beta M(\beta, h_{\beta,\tau^0})$ exist and are continuous at β_0 and $\partial_\beta M(\beta_0, h_0)$ is full column rank. Also by Assumptions 3, 4 and 6 ,proposition 2 and DKW inequality, (2.4) can be satisfied

Next, Condition (2.3) of Theorem 2 can be satisfied by proving the following lemma.

Lemma 5. *For all $\beta \in \mathcal{B}$ and suppose $\beta_0 \in \mathcal{B}$ satisfies $M(\beta_0, h_0) = 0$ and $\beta - \beta_0 = o_p(1)$, the pathwise derivative $\partial_h M(\beta, h_{\beta,\tau^0})$ of $M(\beta, h_{\beta,\tau^0})$ exists in all directions $(h - h_{\beta,\tau^0}) \in \mathcal{H}$. And it satisfies the following two conditions: (1) $\|M(\beta, h) - M(\beta, h_{\beta,\tau^0}) - \partial_h M(\beta, h_{\beta,\tau^0})(h - h_{\beta,\tau^0})\| \leq c\|h - h_{\beta,\tau^0}\|^2_{\mathcal{H}}$. (2) $\|\partial_h M(\beta, h_{\beta,\tau^0})(h - h_{\beta,\tau^0}) - \partial_h M(\beta_0, h_0)(h - h_0)\| \leq o(1)\delta_n$ for each $(\beta, h) \in \Theta_{\delta_n} \times \mathcal{H}_{\delta_n}$ where $\Theta_{\delta_n} = \{\beta \in \mathcal{B} \colon \|\beta - \beta_0\| \leq \delta_n\}$ and $\mathcal{H}_{\theta_n} = \{h \in \mathcal{H} \colon \|h - h_0\| \leq \delta_n\}$.*

Proof. First, it is very easy to prove that the pathwise derivatives of $M(\beta, h_{\beta,\tau^0})$ exists by interchangeability of expectation and differentiation and existence of $\lim_{\epsilon \to 0} \frac{M(\beta, h_0 + \epsilon(h - h_0) - M(\beta, h_0)}{\epsilon}$ Since interchangeability can be proved by Assumptions 3 and 4, we can calculate the derivative by using the definition of pathwise derivative.

$$
\begin{aligned}
\partial_h M(\beta, h_{\beta,\tau^0}) &= E\left[\partial_h \frac{[\partial_\beta G(\tau^0, X_i^T\beta)](Y_i - G(\tau^0, X_i^T\beta))}{G(\tau^0, X_i^T\beta)(1 - G(\tau^0, X_i^T\beta))}\right] \\
&= E\left[\frac{(Y_i - G(\tau^0, X_i^T\beta))}{G(\tau^0, X_i^T\beta)(1 - G(\tau^0, X_i^T\beta))} - \frac{\partial_\beta G(\tau^0, X_i^T\beta)}{G(\tau^0, X_i^T\beta)(1 - G(\tau^0, X_i^T\beta))}\right. \\
&- \frac{\partial_\beta G(\tau^0, X_i^T\beta)\partial_\tau G(\tau^0, X_i^T\beta)}{G(\tau^0, X_i^T\beta)(1 - G(\tau^0, X_i^T\beta))} - \frac{[\partial_\beta G(\tau^0, X_i^T\beta)](Y_i - G(\tau^0, X_i^T\beta))(1 - 2G(\tau^0, X_i^T\beta))}{G(\tau^0, X_i^T\beta)^2(1 - G(\tau^0, X_i^T\beta))^2} \\
&- [\partial_\beta G(\tau^0, X_i^T\beta)]\frac{(Y_i - G(\tau^0, X_i^T\beta))(1 - 2G(\tau^0, X_i^T\beta))\partial_\tau G(\tau^0, X_i^T\beta)}{G(\tau^0, X_i^T\beta)^2(1 - G(\tau^0, X_i^T\beta))^2} \\
&\left. + [\partial_{\beta,\tau} G(\tau^0, X_i^T\beta)]\frac{(Y_i - G(\tau^0, X_i^T\beta))}{G(\tau^0, X_i^T\beta)(1 - G(\tau^0, X_i^T\beta))}\right]
\end{aligned}
$$

By the Law of Iterated Expectations, $\partial_h M(\beta_0, h_0) = E\left[-\frac{\partial_\beta G(\tau^0, X_i^T\beta_0)}{G(\tau^0, X_i^T\beta_0)(1-G(\tau^0, X_i^T\beta_0))} - \frac{\partial_\beta G(\tau^0, X_i^T\beta_0)\partial_\tau G(\tau^0, X_i^T\beta_0)}{G(\tau^0, X_i^T\beta_0)(1-G(\tau^0, X_i^T\beta_0))}\right]$. By Assumptions 3, 4 and 6, we can show that total differential of $M(\beta, h)$ at h_{β,τ^0} will be equal to $\partial_h M(\beta, h_{\beta,\tau^0})(h - h_{\beta,\tau^0})$, thus (1) will be holds. For (2), by compactness mentioned in Assumption 3 and uniform boundedness and differentiability in Assumptions 4 and 6, we can show that $\|\partial_h M(\beta, h_{\beta,\tau^0})(h - h_{\beta,\tau^0}) - \partial_h M(\beta_0, h_0)(h - h_0)\| \leq \|[\partial_h M(\beta, h_{\beta,\tau^0}) - \partial_h M(\beta_0, h_0)]\|\delta_n \leq o_p(1)\delta_n \leq o_p(1)\delta_n$. Thus (2) holds. $\square$

For condition (2.5') and suppose $M(\beta_0, h_0) = 0$, we can rewrite it as follows:

$$\sup_{\|\beta-\beta_0\|\leq\delta_n, \|h-h^0\|_{\mathcal{H}}\leq\delta_n} \|M_n(\beta, h) - M(\beta, h) - (M_n(\beta_0, h_0) - M(\beta_0, h_0))\|_\infty \leq o_p(n^{-1/2})$$

We first define $m(Y_i, X_i, \beta, h) = \frac{[\partial_\beta \hat{G}(\hat{\tau}, X_i^T\beta)]\left(Y_i - \hat{G}(\hat{\tau}, X_i^T\beta)\right)}{\hat{G}(\hat{\tau}, X_i^T\beta)\left(1-\hat{G}(\hat{\tau}, X_i^T\beta)\right)}$. By showing that a class of functions $\mathcal{M} = \{m(Y, X, \beta, h)\colon (y, x, \beta, h) \in \mathcal{YXBH}\}$ is a Donsker class, we can obtain that the $m \in \mathcal{M}$ is stochastic equicontinuity by showing the entropy condition can be satisfied. Then the condition above will hold. By Assumptions 3, 4 and 6, we know that Cartesian product $\mathcal{YXBH}$ is compact and $m(Y, X, \beta, h)$ is point-wise Lipschitz continuous with respect to β and h. From theorem 2.7.11 in Van Der Vaart and Wellner (1996), the bracketing number of $\mathcal{M}$ will be bounded by the covering number of $\mathcal{BH}$ such as $N(\epsilon, \mathcal{M}, \|\cdot\|_{\mathcal{M}}) \leq N(\frac{\epsilon}{L}, \mathcal{BH}, \|\cdot\|_{\mathcal{BH}})$ where L can be set as the $sup_{\mathcal{X}}\|\partial m(\cdot)\|$. By Assumptions 4 and 6, we show that h_1, h_2 and h_3 are Donsker classes by referring to 19.6, 19.9 and 19.20 in Van der Vaart (2000), we show that $m(\cdot)$ belongs to a Donsker classes, which implies that $\int_0^\delta \sqrt{logN(\epsilon, \mathcal{M}, \|\cdot\|_{\mathcal{M}})}d\epsilon \leq \infty$.

Finally we will prove the condition (2.6) by showing the following lemma holds.

$$\begin{aligned}\partial_h M(\beta_0, h_0)(\hat{h} - h_0) = E\Bigg[&-\frac{\partial_\beta G(\tau^0, X_i^T\beta_0)}{G(\tau^0, X_i^T\beta_0)(1 - G(\tau^0, X_i^T\beta_0))}\left(\hat{G}(\hat{\tau}, X_i^T\beta_0) - G(\tau^0, X_i^T\beta_0)\right) \\ &- \frac{\partial_\beta G(\tau^0, X_i^T\beta_0)\partial_\tau G(\tau^0, X_i^T\beta_0)}{G(\tau^0, X_i^T\beta_0)(1 - G(\tau^0, X_i^T\beta_0))}(\hat{\tau} - \tau^0)\Bigg]\end{aligned}$$

For the first part, we define $\alpha(z) = \frac{E(\partial_\beta G(\tau^0,z)|z)}{G(\tau^0,z)(1-G(\tau^0,z))}$ and have that

$$
\begin{aligned}
&\int \alpha(z)[\hat{G}(\hat{\tau}, z) - G(\tau^0, z)]f(z,\tau_0)/f(\tau_0)dz \\
=&\int \alpha(z)\left[\hat{G}(\hat{\tau}, z) - \hat{G}(\tau^0, z) + \hat{G}(\tau^0, z) - G(\tau^0, z)\right] f(z,\tau_0)/f(\tau_0)dz \\
=&\int \alpha(z)\left[\frac{\frac{1}{nb}\sum_i 1\{Y_i^{*c} \le \hat{\tau}_1, Y_i^{*p} \le \hat{\tau}_0\}K(\frac{z_i - z}{b})}{\hat{f}(z,\hat{\tau}_0)} - \frac{\frac{1}{nb}\sum_i 1\{Y_i^{*c} \le \tau_1, Y_i^{*p} \le \tau_0\}K(\frac{z_i - z}{b})}{\hat{f}(z,\tau_0)}\right] f(z,\tau_0)/f(\tau_0)dz \\
&+ \int \alpha(z)\left[\frac{\frac{1}{nb}\sum_i 1\{Y_i^{*c} \le \tau_1, Y_i^{*p} \le \tau_0\}K(\frac{z_i - z}{b})}{\hat{f}(z,\tau_0)} - G(\tau^0, z)\right] f(z,\tau_0)/f(\tau_0)dz \\
=&\int \alpha(z)\left[\frac{\bar{m}(\hat{\tau}, z)}{\hat{f}(z,\hat{\tau}_0)} - \frac{\bar{m}(\tau, z)}{\hat{f}(z,\tau_0)}\right] f(z,\tau_0)/f(\tau_0)dz + \int \alpha(z)\left[\frac{\bar{m}(\tau, z)}{\hat{f}(z,\tau_0)} - G(\tau^0, z)\right] f(z,\tau_0)/f(\tau_0)dz \\
=&\int \alpha(z)\left[\frac{\bar{m}(\hat{\tau}, z) - \bar{m}(\tau, z)}{f(z,\tau_0)}\right] f(z,\tau_0)/f(\tau_0)dz + \int \alpha(z)\left[\frac{G(\tau,z)(f(z,\tau_0) - \hat{f}(z,\hat{\tau}_0))}{f(z,\tau_0)}\right] f(z,\tau_0)/f(\tau_0)dz \\
&+ \int \alpha(z)\left[\frac{\bar{m}(\tau, z) - g_{\tau,z}(\tau, z)}{f(z,\tau_0)}\right] f(z,\tau_0)/f(\tau_0)dz + o_p(n^{-1/2}) \\
&= D_1 + D_2 + D_3
\end{aligned}
$$

where b is the bandwidth. For D_1, we can rewrite as follows:

$$
\begin{aligned}
&\int \alpha(z)\left[\frac{\bar{m}(\hat{\tau}, z) - \bar{m}(\tau, z)}{f(z,\tau_0)}\right] f(z,\tau_0)/f(\tau_0)dz \\
=&\int \alpha(z)\left[\frac{\bar{m}(\hat{\tau}, z) - \bar{m}^*(\hat{\tau}, z) - (\bar{m}(\tau, z) - \bar{m}^*(\tau, z))}{f(\tau_0)}\right] dz + \int \alpha(z)\left[\frac{\bar{m}^*(\hat{\tau}, z) - \bar{m}^*(\tau, z)}{f(\tau_0)}\right] dz \\
=&\int \frac{\alpha(z)}{f(\tau_0)}\left[\phi_0(\tau_0,\tau_1,z)(\hat{\tau}_0 - \tau_0) + \phi_1(\tau_0,\tau_1,z)(\hat{\tau}_1 - \tau_1)\right] f(z)dz
\end{aligned}
$$

where $\phi_0(\tau_0,\tau_1,z) = \frac{\partial_{\tau_0} G(\tau,z) f(z,\tau_0) + \partial_{\tau_0} f(z,\tau_0) G(\tau,z)}{f(z)}$ and $\phi_1(\tau_0,\tau_1,z) = \frac{\partial_{\tau_1} G(\tau,z) f(z,\tau_0)}{f(z)}$. The first term in D_1 will converge to zero by stochastic equicontinuity which will be satisfied by Assumptions 4 and 6. Next we move to D_2, the same process for D_1 can be applied to D_2.

$$
\begin{aligned}
&\int \alpha(z)\left[\frac{G(\tau,z)(f(z,\tau_0) - \hat{f}(z,\hat{\tau}_0))}{f(z,\tau_0)}\right] f(z,\tau_0)/f(\tau_0)dz \\
=&-\int \frac{\alpha(z)}{f(\tau_0)}\left[G(\tau,z)\left(\hat{f}(z,\hat{\tau}_0) - \hat{f}(z,\tau_0) + (\hat{f}(z,\tau_0) - f(z,\tau_0))\right)\right] dz \\
=&-\int \frac{\alpha(z)G(\tau,z)}{f(\tau_0)}\left[\phi_0^p(\tau_0,z)(\hat{\tau}_0 - \tau_0)\right] f(z)dz - \int \frac{\alpha(z)}{f(\tau_0)}\left[G(\tau,z)\left(\hat{f}(z,\tau_0) - f(z,\tau_0)\right)\right] dz \\
=&-\int \frac{\alpha(z)G(\tau,z)}{f(\tau_0)}\left[\phi_0^p(\tau_0,z)(\hat{\tau}_0 - \tau_0)\right] f(z)dz - \left[\frac{1}{n}\sum_i \frac{\alpha(z_i)G(\tau,z_i)}{f(\tau_0)} - E_{z|\tau_0}\left(G(\tau,z)\alpha(z)\right)\right]
\end{aligned}
$$

where $\phi_0^p(\tau_0, z)$ is the derivative of the distribution of y conditional on z and equal to $\frac{\partial_{\tau_0} f(z,\tau_0)}{f(z)}$. $1\{Y_i^{*p} \leq \tau_0\}G(\tau, z_i) = G(\tau, z_i)$ since $G(\tau, z_i)$ has already defined under the condition $1\{Y_i^{*p} \leq \tau_0\}$. For D_3, we can rewrite as

$$
\begin{aligned}
&\int \alpha(z)\left[\frac{\bar{m}(\tau, z) - \phi(\tau_0, \tau_1, z) f(z)}{f(z, \tau_0)}\right] f(z, \tau_0)/f(\tau_0) dz + o_p(n^{-1/2}) \\
=&\frac{1}{n}\sum_i \frac{\alpha(z_i)}{f(\tau_0)} 1\{Y_i^{*c} \leq \tau_1, Y_i^{*p} \leq \tau_0\} - E_{z|\tau_0}\left(G(\tau, z)\alpha(z)\right) + o_p(n^{-1/2})
\end{aligned}
$$

For the second part J_2, we can obtain:

$$
E\left[\frac{\partial_\beta G(\tau^0, X_i^T\beta_0)\partial_{\tau_0} G(\tau^0, X_i^T\beta_0)}{G(\tau^0, X_i^T\beta_0)(1 - G(\tau^0, X_i^T\beta_0))}\right](\hat{\tau}_0 - \tau_0) + E\left[\frac{\partial_\beta G(\tau^0, X_i^T\beta_0)\partial_{\tau_1} G(\tau^0, X_i^T\beta_0)}{G(\tau^0, X_i^T\beta_0)(1 - G(\tau^0, X_i^T\beta_0))}\right](\hat{\tau}_1 - \tau_1)
$$

Putting together the terms above, we can write $M_n(\beta_0, \tau^0) + \partial_\beta M(\beta_0, \tau^0)(\hat{h}_{\beta_0,\hat{\tau}} - h^0)$ as:

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \frac{\left[\partial_\beta G(\tau^0, X_i^T\beta_0) - E(\partial_\beta G(\tau^0, X_i^T\beta_0)|X_i)\right]\left(1\{Y_i^{*c} \leq \tau_1, Y_i^{*p} \leq \tau_0\} - G(\tau^0, X_i^T\beta_0)\right)}{f(\tau_0) G(\tau^0, X_i^T\beta_0)\left(1 - G(\tau^0, X_i^T\beta_0)\right)} \\
&+ \frac{1}{n}\sum_{i=1}^n (\Psi_0\psi_{i0} + \Psi_1\psi_{i1})
\end{aligned}
$$

where $(\hat{\tau}_j - \tau_j) = \frac{1}{n}\sum_{i=1}^n \psi_{ij}$ and $\Psi_j = E\left[\frac{\left(\partial_\beta G(\tau^0, X_i^T\beta_0) - E(\partial_\beta G(\tau^0, X^T\beta_0)|z)\right)\partial_{\tau_0} G(\tau^0, X_i^T\beta_0)}{G(\tau^0, X_i^T\beta_0)(1 - G(\tau^0, X_i^T\beta_0))}\right]\psi_{ij}$ for j =1,2.

The results of condition (2.6) can be followed by using standard central limit theorem and lemma 6. Then by the Theorem 2 in Chen, Linton, and Van Keilegom (2003), we can show that the asymptotic normality can be proved.

Lemma 6. *Under Assumption 5,* $\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, V_\tau)$

Proof. Define the distribution of i.i.d sample $Y_1, ..., Y_n$ as $F_{Y^c}(y^c)$ and the sample distribution $\hat{F}_Y = n^{-1}\sum_{i=1}^n H_i$ where $H_i = \mathbf{1}\{Y_i \leq y\}$ follows an Bernoulli distribution. By central limit theorem of empirical distribution, we can get: $\sqrt{n}(\hat{F}_Y - F_Y) \xrightarrow{d} N(0, V_F)$ where $V_F = F_Y(y)(1 - F_Y(y))$. Then we can get the asymptotic distribution of $\hat{\tau} - \tau$ by using the Delta method: $\sqrt{n}(\hat{\tau} - \tau) = \sqrt{n}(F^{-1}(\hat{F}_Y) - F^{-1}(F_Y)) = \frac{\sqrt{n}(\hat{F}_Y - F_Y)}{f(F^{-1}(q))}$ where q is the qth quantile. Then we can get

$$\sqrt{n}(\hat{\tau}-\tau) = \frac{\sqrt{n}(\hat{q}-q)}{f(F^{-1}(q))} + o_p(1)$$
$$\xrightarrow{d} N(0, V_\tau) \tag{9}$$

where $V_\tau = \frac{V_F}{f(\tau)^2} = \frac{q(1-q)}{f(\tau)^2}$ and $f(\tau)$ is the value of the probability density at the q-th quantile.

□

Proposition 3. *Under the conditions of Montiel Olea and Plagborg-Møller (2019), Denote $\hat{u}^* \in \mathbb{R}^p$ as a random vector whose distribution conditional on the data is denoted as $\hat{P}$. Let $\hat{P}_B$ denote the distribution of $\sqrt{n}(\hat{u}^* - \hat{u})$ conditional on the data. Let P denote the limit distribution of $\sqrt{n}(\hat{u} - u)$. If $\rho(\hat{P}_B - P) \xrightarrow{p} 0$ as $n \to \infty$ where $\rho(\cdot, \cdot)$ denotes metric that metricizes weak convergence of probability measures on $\mathbb{R}^p$.*

(i) Assume for each j = 1,...,k, there exists a random variable $\hat{\sigma}_j^*$ such that $\sqrt{n}\hat{\sigma}_j^* \xrightarrow{p} \Sigma_{jj}^{1/2}$. Let $\hat{q}_{1-\alpha}$ denote the $1-\alpha$ quantile of the distribution of $\max_j(\hat{\sigma}_j^*)^{-1}|h_j(\hat{u}^*) - h_j(\hat{u})|$ conditional on the data. Then

$$\hat{q}_{1-\alpha} \xrightarrow{p} q_{1-\alpha}(\Sigma)$$

(ii) Denote the ζ quantile of $h_j(\hat{u}^*)$ conditional on the data as $\hat{Q}_{j,\zeta}$. Define ζ_{max} as the largest value of $\zeta \in [0, 1/2]$ such that $\hat{P}(h_j(\hat{u}^*) \in \times_{j=1}^k [\hat{Q}_{j,\zeta}, \hat{Q}_{j,1-\zeta}]) \geq 1-\alpha$, conditional on the data. Let $\Phi(\cdot)$ denote the standard normal CDF. Then

$$\zeta_{max} \xrightarrow{p} \zeta^* \equiv \Phi(-q_{1-\alpha}(\Sigma))$$

(iii) Under the same conditions as in (ii), we have for any j=1,...,k,

$$\hat{Q}_{j,\zeta} = \hat{\theta}_j - \hat{\sigma}_j q_{1-\alpha}(\Sigma) + o_p(n^{-1/2}),$$

$$\hat{Q}_{j,1-\zeta} = \hat{\theta}_j + \hat{\sigma}_j q_{1-\alpha}(\Sigma) + o_p(n^{-1/2})$$

Proof. The Proposition 3 in the appendix is the Proposition 3 in Montiel Olea and Plagborg-Møller (2019). We will briefly discuss the proof of the Proposition 3. By the conditions of

Theorem 3, the Assumption 1 of Montiel Olea and Plagborg-Møller (2019) can be satisfied by letting the function $\theta = h(u) = u$. Then by the results of Theorem 3, the assumption $\rho(\hat{P}_B - P) \xrightarrow{p} 0$ can be satisfied. According to the Appendix B.4.9 inMontiel Olea and Plagborg-Møller (2019), we can show that the results $(i), (ii)$ and (iii) hold. □

Proof of Proposition 1. In order to prove this proposition, we have to show that $|\partial_{x_{di}}\hat{G}(\hat{\tau}, X^T\hat{\beta}) - \partial_{x_{di}}G(\tau, X^T\beta)| \leq |\partial_{x_{di}}\hat{G}(\hat{\tau}, X^T\hat{\beta}) - \partial_{x_{di}}G(\hat{\tau}, X^T\hat{\beta})| + |\partial_{x_{di}}G(\tau, X^T\hat{\beta}) - \partial_{x_{di}}G(\tau, X^T\beta)| + |\partial_{x_{di}}G(\hat{\tau}, X^T\beta) - \partial_{x_{di}}G(\tau, X^T\beta)| + |\partial_{x_{di}}G(\hat{\tau}, X^T\hat{\beta}) - \partial_{x_{di}}G(\tau, X^T\hat{\beta}) - (\partial_{x_{di}}G(\hat{\tau}, X^T\beta) - \partial_{x_{di}}G(\tau, X^T\beta))| = o_p(1)$ as n goes to infinity. According to the lemma 2 in Klein and Spady, we can get that the first term on the right of the inequality will converge to zero. By Assumptions 3 and 4, the second and third term can be proved to converge to zero in probability by showing that both terms are satisfying Lipschitz condition and consistency of the estimators of τ and β also holds. For the last term, we show that the stochastic equicontinuity holds by showing that $V_n(\hat{\beta}) = \partial_{x_{di}}G(\hat{\tau}, X^T\hat{\beta}) - \partial_{x_{di}}G(\tau, X^T\hat{\beta})$ belongs to a Donsker class by referring to 19.9 in Van der Vaart (2000). □