

DISCUSSION PAPER SERIES

IZA DP No. 10473

**Estimation and Inference for Actual and
Counterfactual Growth Incidence Curves**

Francisco H. G. Ferreira
Sergio Firpo
Antonio F. Galvao

JANUARY 2017

DISCUSSION PAPER SERIES

IZA DP No. 10473

Estimation and Inference for Actual and Counterfactual Growth Incidence Curves

Francisco H. G. Ferreira

The World Bank and IZA

Sergio Firpo

Inspira and IZA

Antonio F. Galvao

University of Iowa

JANUARY 2017

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Estimation and Inference for Actual and Counterfactual Growth Incidence Curves*

Different episodes of economic growth display widely varying distributional characteristics, both across countries and over time. Growth is sometimes accompanied by rising and sometimes by falling inequality. Applied economists have come to rely on the Growth Incidence Curve, which gives the quantile-specific rate of income growth over a certain period, to describe and analyze the incidence of economic growth. This paper discusses the identification conditions, and develops estimation and inference procedures for both actual and counterfactual growth incidence curves, based on general functions of the quantile potential outcome process over the space of quantiles. The paper establishes the limiting null distribution of the test statistics of interest for those general functions, and proposes resampling methods to implement inference in practice. The proposed methods are illustrated by a comparison of the growth processes in the United States and Brazil during 1995-2007. Although growth in the average real wage was disappointing in both countries, the distribution of that growth was markedly different. In the United States, wage growth was mediocre for the bottom 80 percent of the sample, but much more rapid for the top 20 percent. In Brazil, conversely, wage growth was rapid below the median, and negative at the top. As a result, inequality rose in the United States and fell markedly in Brazil.

JEL Classification: C14, C21, D31, I32

Keywords: growth incidence curves, potential outcomes, inference, quantile process

Corresponding author:

Francisco H. G. Ferreira
World Bank
1818 H Street, NW
Washington, DC 20433
USA

E-mail: fferreira@worldbank.org

* The authors would like to express their appreciation to Matias Cattaneo, Yu-Chin Hsu, David Kaplan, Ying-Ying Lee, Zhongjun Qu, Alexandre Poirier, Liang Wang and participants at the 2015 meeting of the Midwest Econometric Group for useful comments and discussions regarding this paper. Vitor Possebom provided excellent research assistance. Computer programs to replicate the numerical analyses are available from the authors. All the remaining errors are ours.

1 Introduction

Growth episodes have displayed widely different distributional characteristics across countries and over time. The same rate of growth in average incomes has been accompanied by rising inequality in some cases, and by falling inequality in others. A large literature on “pro-poor growth” and, more generally, on the incidence of economic growth processes has developed, and attracted attention among both researchers and policymakers.

Over time, this literature has come to rely heavily on the Growth Incidence Curve (*GIC*), which describes the rate of income growth at each quantile $\tau \in (0, 1)$ of the (anonymous) distribution (Ravallion and Chen (2003)). It has been used to compare the distributional characteristics of growth processes both across countries and over time (see, e.g. Besley and Cord (2007)). It has also been shown to underlie changes in certain widely-used classes of poverty and inequality measures, which can be formally expressed as functionals of the *GIC* (Ferreira (2012)).

Growth incidence curves have also featured in a long-standing literature that uses counterfactual income distributions to decompose changes (or differences) in inequality and poverty over time (or between countries), and to attribute such changes to different factors such as, for example, changes in worker characteristics or in the returns to those characteristics. The original contributions to this literature, including Juhn, Murphy, and Pierce (1993), Dinardo, Fortin, and Lemieux (1996) and Donald, Green, and Paarsch (2000), predate the Ravallion and Chen (2003) article that introduced the term *GIC*, and hence do not use it. Yet, each of those papers sought to account for differences across entire wage or income distributions – which can be formally expressed as *GICs* – using counterfactual distributions. Ferreira (2012) defines counterfactual growth incidence curves as functionals of counterfactual distributions, and establishes the link to this earlier literature on distributional change.

Despite their conceptual importance and widespread practical use, however, formal conditions for identification and inference using growth incidence curves – actual or counterfactual – have not been established. In this paper, we rely on the formal analogy between distributional change and treatment heterogeneity to fill that gap. More specifically, we write both actual and counterfactual *GICs* in terms of vectors of potential outcomes (Rubin (1977)), and then apply suitable variants of a number of results from the literature on quantile treatment effects to formally establish the conditions for identification of the *GIC*. Specifically, we adapt the identification results in Firpo (2007), where the relevant identification restric-

tion is the ignorability assumption.¹ In our context, it implies that the income distributions that we observe in two different time periods are generated by two group of factors only: observable components whose distributions may vary over time, and unobservable components whose conditional distributions given observables are fixed over time.

We then propose a simple three-step semiparametric estimator for both actual and counterfactual growth incidence curves, which relies on established sample re-weighting and quantile regression techniques. In the first step, a nonparametric estimator of the propensity score is used, and weights are computed. In our setup, the propensity score is computed by pooling the repeated cross-section data for initial and end periods and calculating the probability of being observed at the final period, given covariates. In the second step, one obtains properly weighted quantiles of the outcome from a simple weighted quantile regression. The third step is the computation of the *GIC* as a function of the vector of quantiles of weighted outcome distributions.² When applied to counterfactual *GICs*, this procedure has the added advantage that it requires no assumption on the structural relationships between income and its covariates, as was the case with most of the previous literature.

We establish the asymptotic properties of these estimators, propose suitable test statistics, and discuss inference procedures in practice. For practical inference we compute critical values using resampling methods. We provide sufficient conditions and show the theoretical validity of a bootstrap approach. Moreover, we discuss in detail an algorithm for its practical implementation. We also discuss computation of critical values through a subsampling method.

The main technical contributions of the paper are as follows. The first is to develop practical statistical inference procedures for the *GIC*. This enables researchers to conduct estimation and inference for the *GIC* over the entire set of quantiles. Secondly, we can easily extend our results to general functionals of the vector of quantiles of potential outcomes and not only the one that yields the *GIC*, which allows us to develop testing procedures for general hypotheses involving these functionals.³ An additional by-product contribution of

¹This condition has been employed widely in the distributional treatment effect literature. See not only Firpo (2007), but also, among others, Flores (2007), Cattaneo (2010), and Galvao and Wang (2015).

²A natural extension of our method – not pursued in this paper – would be to implement a fourth step, which would involve estimation and inference of real-valued functionals of the *GIC* process, such as poverty and income inequality growth.

³The theoretical results derived in this paper can be applied to other functionals of the quantiles of potential outcomes processes. For instance, the quantile treatment effects in Firpo (2007), and the Makarov bounds for the quantiles of the distribution of treatment effects discussed in Fan and Park (2010), although following a more elaborate formula, are also functionals of the quantiles of the potential outcomes. In general, our final estimator can be seen as a plug-in estimator of the functional using the estimated quantiles

this paper is to establish the asymptotic properties of the estimator of the vector of quantiles of weighted outcome distributions for the quantile process, namely, uniform consistency and weak convergence. The provision of uniform results over the set of quantiles is a necessary condition to establish the results for the testing procedures. We also show that the estimator is uniformly efficient, as the asymptotic variance of the estimator coincides with the semiparametric efficiency bound.

These contributions are closely related to the literature on quantile treatment effects, which is a particular functional of the vector formed by the quantiles of the potential outcomes. That literature started with Doksum (1974) and Lehmann (1974) and has expanded recently (see, e.g., Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2005), Bitler, Gelbach, and Hoynes (2006), Firpo (2007), Cattaneo (2010), Donald and Hsu (2014), Galvao and Wang (2015), and Firpo and Pinto (2015)).⁴

We illustrate the proposed procedure by comparing actual and counterfactual growth incidence curves (for real hourly wages) for the two largest countries in the Western Hemisphere, namely the United States and Brazil, in the twelve years prior to the onset of the last great financial crisis: 1995-2007. Although growth rates in average wages were disappointing in both countries (especially in Brazil), there were substantial differences in inequality dynamics. The *GIC* for the US was flat until approximately the 8th decile, and sharply upward-sloping over the top quintile. In Brazil conversely, the *GIC* peaked around the first quintile, and was downward sloping thereafter. As a result wage inequality rose sharply in the US and declined in Brazil.

We use counterfactual *GICs* to examine whether these changes were driven primarily by the composition of the labor force - in terms of observed worker characteristics such as gender, age, and education - or by changes in the broader structure of the economy. In both countries, we find that increases in worker age (and thus potential experience) and education contributed to income growth in a roughly equiproportional manner. Changes in inequality were driven almost entirely by changes in economic structure.

The remainder of the paper is organized as follows. Section 2 defines the *GIC*. Section 3 presents the econometric results, describes the three-step estimator, establishes the asymptotic properties of the estimator, discusses inference for the quantile process, and its of potential outcomes.

⁴The results of this paper are also related to those on inference on the quantile process. See, e.g., Belloni, Chernozhukov, and Fernandez-Val (2011), and Qu and Yoon (2015) for the nonparametric case; Gutenbrunner and Jureckova (1992), Koenker and Machado (1999), Koenker and Xiao (2002), Chernozhukov and Fernandez-Val (2005), and Angrist, Chernozhukov, and Fernandez-Val (2006) for the parametric case.

practical implementation. The empirical application to the US and Brazil is presented in Section 4. Section 5 concludes. We relegate the proofs of the results to the Appendix.

2 Growth incidence curves: Actual and counterfactual

In this section we formally define the growth incidence curve (*GIC*), which was originally introduced by Ravallion and Chen (2003). Let Y be the outcome variable of interest, say an indicator of economic welfare such as income. There are two time periods, 0 and 1. Let us say that an individual observation taken at time 1 belongs to group A , ie, $G = A$. An observation taken at time 0 belongs to group B , or $G = B$. Assume that income is continuously distributed over the population of interest, and denote its cumulative distribution function (CDF) at time t as $F_{Y|T}(\cdot|t)$. The income level at the τ -th quantile for groups A and B are given by, respectively, the inverse of the CDF, $q_A(\tau) = F_{Y|T}^{-1}(\tau|1)$ and $q_B(\tau) = F_{Y|T}^{-1}(\tau|0)$. Then, the instantaneous *GIC* at a given time t and quantile τ can be represented as $\frac{dF_{Y|T}^{-1}(\tau|t)/dt}{F_{Y|T}^{-1}(\tau|t)}$. In discrete time, the income growth rate for a given quantile τ between two time periods, 0 and 1, can then be written as

$$GIC_Y(\tau) = \frac{q_A(\tau) - q_B(\tau)}{q_B(\tau)}.$$

Motivated by the importance of the *GIC* for the economic analysis of social welfare, this paper develops estimation and inference procedures for the $GIC(\tau)$, which is calculated as the difference of quantiles in time periods 1 and 0 over the quantile in time zero, for the entire set of quantiles $\tau \in (0, 1)$. We assume availability of a random sample of size n from the joint distribution of (Y, T, \mathbf{X}) , where Y is the income, T is a time dummy variable that equals 1 at period $T = 1$, and \mathbf{X} is a vector of length d of covariates. We could have represented the data equivalently as (Y, G, \mathbf{X}) .

The covariates enable us to learn how changes in their joint distribution affect growth and inequality. For an individual i in our sample, if $G_i = A$ we observe $Y_i(1)$, otherwise $G_i = B$ and we observe $Y_i(0)$, where $Y_i(1)$ is what individual i 's outcome would be were she observed at time $T = 1$, and $Y_i(0)$ is what individual i 's outcome would be were she observed at time $T = 0$. Borrowing from the treatment effect literature, we call $Y(1)$ and $Y(0)$ ‘potential outcomes’; we say that individual i is ‘treated’ if she is observed at period 1 or group A , and ‘untreated’ if observed at period 0 or group B . We may refer to T as the

‘treatment assignment dummy’ or, more accurately, ‘time assignment dummy’. Thus, the observed outcome is $Y = (Y(1) - Y(0))T + Y(0)$.

Writing the problem in terms of potential outcomes is useful because it allows us to easily write both actual and counterfactual distributions. For example, the actual outcome distribution for those individuals from group B , that is, those who were observed at time 0, is $F_{Y(0)|T}(\cdot|0)$ and the actual outcome distribution for those individuals from group A , that is, those who were observed at time 1, is $F_{Y(1)|T}(\cdot|1)$. The counterfactual outcome distribution for those individuals who were observed at time 0, were they observed at time 1, is $F_{Y(1)|T}(\cdot|0)$ and the counterfactual outcome distribution for those individuals who were observed at time 1, were they observed at time 0, is $F_{Y(0)|T}(\cdot|1)$.

Let τ be a real number in $\mathcal{T} \subset (0, 1)$ and $t = 0, 1$. Let $q_{At}(\tau)$ be $\inf_q \Pr[Y(t) \leq q | T = 1] \geq \tau$, or the τ th quantile of $F_{Y(t)|T}(\cdot|1)$, which is the distribution function of $Y(t)$ for the subpopulation A . For the B subpopulation, let $q_{Bt}(\tau)$ be $\inf_q \Pr[Y(t) \leq q | T = 0] \geq \tau$, or the τ th quantile of $F_{Y(t)|T}(\cdot|0)$. For both subpopulations, those distribution functions share the same support, which is $\mathcal{Y}_t \subset \mathbb{R}$. Let us also define

$$Q_A(\tau, \tau') := \begin{bmatrix} q_{A1}(\tau) \\ q_{A0}(\tau') \end{bmatrix}, \text{ and } Q_B(\tau, \tau') := \begin{bmatrix} q_{B1}(\tau) \\ q_{B0}(\tau') \end{bmatrix}.$$

Thus, the *GIC* can be derived from the previous variables as the growth rate of income at the τ th quantile between periods 0 and 1. We first define the observed or actual *GIC* as

$$GIC(\tau) := \frac{q_{A1}(\tau) - q_{B0}(\tau)}{q_{B0}(\tau)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{matrix} Q_A(\tau, \tau) \\ Q_B(\tau, \tau) \end{matrix} - 1. \quad (1)$$

The graphical depiction of *GIC*, as proposed in Ravallion and Chen (2003), is obtained by letting τ vary from zero to one and plotting the corresponding values of *GIC* against the quantiles τ . The quantiles involved in the computation of equation (1) are based on the ranking of individuals in each distribution of interest. Therefore, unless the individual i keeps her ranking over time, *GIC* will not be an appropriate tool to infer individual gains over time. This is a consequence of the veil of ignorance (anonymity) shrouding the comparison of the two distributions (see Essama-Nssah, Paul, and Bassole (2013)).

The interpretation of the graphical depiction of *GIC* is simple. If the *GIC* is a decreasing function for all τ in its domain of definition, then all inequality measures that respect the

Pigou-Dalton principle of transfers and scale invariance will indicate a fall in inequality over time. If instead, the GIC is an increasing function of τ , then the same measures will register an increase in inequality (Ravallion and Chen (2003)). When no relative inequality measure changes over time, then the GIC will present a constant growth rate over the process of quantiles τ .

Using our previous notation, we can define GIC^* as the counterfactual GIC . It can be expressed as

$$GIC^*(\tau) := \frac{q_{B1}(\tau) - q_{B0}(\tau)}{q_{B0}(\tau)} = \frac{\begin{bmatrix} 1 & -1 \end{bmatrix} Q_B(\tau, \tau)}{\begin{bmatrix} 0 & 1 \end{bmatrix} Q_B(\tau, \tau)} = \frac{\begin{bmatrix} 1 & 0 \end{bmatrix} Q_B(\tau, \tau)}{\begin{bmatrix} 0 & 1 \end{bmatrix} Q_B(\tau, \tau)} - 1, \quad (2)$$

which is the growth incidence curve for quantile τ if the distribution of associated factors (explanatory variables, or covariates) had remained fixed from period 0 to 1. GIC^* captures only that part of distributional change associated with changes in the conditional distribution $F_{Y(\cdot)|T}$, which we interpret broadly as changes in the structure of the economy.

Comparing GIC with GIC^* allows us to understand whether heterogeneity in economic growth is driven by changes in the joint distribution of observed covariates (\mathbf{X}) that impact income, or is driven by changes in the structure of the economy. For example, if GIC is decreasing in τ but GIC^* is uniform (flat) over τ , the decrease in inequality is driven by changes in the distribution of covariates. This interpretation can be formally obtained by decomposing the $GIC(\tau)$ into two components as following:

$$GIC(\tau) = GIC^*(\tau) + GIC^{**}(\tau) \cdot \frac{q_{B1}(\tau)}{q_{B0}(\tau)},$$

where

$$GIC^{**}(\tau) := \frac{q_{A1}(\tau) - q_{B1}(\tau)}{q_{B1}(\tau)} = \frac{\begin{bmatrix} 1 & 0 \end{bmatrix} Q_A(\tau, \tau)}{\begin{bmatrix} 1 & 0 \end{bmatrix} Q_B(\tau, \tau)} - 1$$

is the growth incidence curve that would have occurred only because of time changes in the distribution of covariates.

We will develop estimation and inference procedures for the $GIC(\tau)$ and $GIC^*(\tau)$ and, more generally, for functionals of the quantile of potential outcomes. In that sense, our theoretical framework provides a flexible method for the practical analysis of the growth incidence curves.

3 The econometric model

In this section we introduce the econometric model, discuss identification, estimation of the parameters of interest, and inference procedures. As previously seen, GIC can be written as a function of the vector of quantiles of potential outcomes. Thus, in this section, we first obtain the results for the latter, and then, for the GIC . **Notation:** Let \mathbb{E} and \mathbb{E} be expectation and sample average, respectively. Let \rightsquigarrow , \xrightarrow{p} , and $\xrightarrow{p^*}$ denote weak convergence, and convergence in probability and in outer probability, respectively. Let $\|g(z)\|_\infty$ denote $\sup_z |g(z)|$ for $z \in \mathcal{Z}$.

3.1 Identification

In order to make our setup comparable with the treatment effects literature, we maintain all definitions and notation as it is commonly used in that framework. Therefore, we have a random sample of size n from the joint distribution of (Y, T, \mathbf{X}) , where Y is the outcome of interest, T is a dummy variable of treatment assignment, and \mathbf{X} is a vector of length d of covariates. For completeness, in this section, we also define $q_t(\tau)$ as $\inf_q \Pr[Y(t) \leq q] \geq \tau$, for $t = 0, 1$, which is the unconditional τ th quantile of $F_{Y(t)}$, the distribution function of $Y(t)$ whose support is $\mathcal{Y}_t \subset \mathbb{R}$.

Now we define the p -score, the conditional probability of being treated (observed at time 1) given \mathbf{X} , as $p(\mathbf{X})$, and the unconditional probability as p . Let $\mathbf{X} \in \mathcal{X} \subset \mathbb{R}^d$.

In what follows, it is also useful to define the function m as: $m(a, b; \tau) = \tau - \mathbf{1}\{a < b\}$. We state assumptions on the general model for identification of the parameters of interest.

I.I For each $\tau \in \mathcal{T}$, $t = 0, 1$, $q_t(\tau)$ uniquely solves $\mathbb{E}[m(Y(t), q_t(\tau); \tau)] = 0$; $q_{A_t}(\tau)$ uniquely solves $\mathbb{E}[m(Y(t), q_{A_t}(\tau); \tau) | T = 1] = 0$; and $q_{B_t}(\tau)$ uniquely solves $\mathbb{E}[m(Y(t), q_{B_t}(\tau); \tau) | T = 0] = 0$.

I.II For all $\tau \in \mathcal{T}$, we have $(Y(1), Y(0)) \perp T | \mathbf{X}$;

I.III For some $c > 0$, $c < p(\mathbf{X}) < 1 - c$, a.e. \mathbf{X} .

Assumptions **I.I–I.III** are standard in the literature, as in Firpo (2007). Condition **I.I** is in general not a sufficient identification condition for $q_t(\tau)$ because $Y(t)$ is not always observable from the data. Therefore, the untestable condition **I.II**, the so-called ignorability assumption, is fundamental. According to condition **I.II**, the assignment to the treatment is

random within subpopulations characterized by \mathbf{X} . This assumption has been used, among others, by Heckman, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), Hirano and Imbens (2004), Firpo (2007). Within the *GIC* framework this assumption implies that conditional on \mathbf{X} , there is a random mechanism that assigns individual i to the exact period that she is observed (either period 0 or 1). In our model the triple (Y, T, \mathbf{X}) is observable, and a random sample of size n can be obtained. Condition **I.III** states that for almost all values of \mathbf{X} , both treatment assignment levels have a positive probability of occurrence.

Under conditions **I.I–I.III** the quantities $q_1(\tau)$, $q_0(\tau)$, $q_{A1}(\tau)$, $q_{A0}(\tau)$, $q_{B1}(\tau)$ and $q_{B0}(\tau)$ are identified from the joint distribution of (Y, T, \mathbf{X}) . These six objects can be written as implicit functions of the observed data. For all $\tau \in \mathcal{T}$,

$$\begin{aligned} \mathbb{E}[w_1(T, \mathbf{X}) m(Y, q_1(\tau); \tau)] &= \mathbb{E}[w_0(T, \mathbf{X}) m(Y, q_0(\tau); \tau)] \\ &= \mathbb{E}[w_{A1}(T, \mathbf{X}) m(Y, q_{A1}(\tau); \tau)] = \mathbb{E}[w_{A0}(T, \mathbf{X}) m(Y, q_{A0}(\tau); \tau)] \\ &= \mathbb{E}[w_{B1}(T, \mathbf{X}) m(Y, q_{B1}(\tau); \tau)] = \mathbb{E}[w_{B0}(T, \mathbf{X}) m(Y, q_{B0}(\tau); \tau)] = 0, \end{aligned}$$

where $w_1(T, \mathbf{X}) = \frac{T}{p(\mathbf{X})}$, $w_0(T, \mathbf{X}) = \frac{1-T}{1-p(\mathbf{X})}$, $w_{A1}(T, \mathbf{X}) = \frac{T}{p}$, $w_{A0}(T, \mathbf{X}) = \left(\frac{1-T}{p}\right) \left(\frac{p(\mathbf{X})}{1-p(\mathbf{X})}\right)$, $w_{B1}(T, \mathbf{X}) = \left(\frac{T}{1-p}\right) \left(\frac{1-p(\mathbf{X})}{p(\mathbf{X})}\right)$ and $w_{B0}(T, \mathbf{X}) = \frac{1-T}{1-p}$. This main identification result follows directly from Lemma 1 in Firpo (2007).

Finally, given that the elements in the vectors $Q(\tau, \tau')$, $Q_A(\tau, \tau')$ and $Q_B(\tau, \tau')$ are identified, since $q_1(\tau)$, $q_0(\tau)$, $q_{A1}(\tau)$, $q_{A0}(\tau)$, $q_{B1}(\tau)$ and $q_{B0}(\tau)$ are identified, it follows from equations (1) and (2) that $GIC(\tau)$ and $GIC^*(\tau)$ are also, respectively, identified from the joint distribution of (Y, T, \mathbf{X}) .

Remark 1. *We note that one can also obtain other functionals of interest based on $Q(\tau, \tau')$, $Q_A(\tau, \tau')$ and $Q_B(\tau, \tau')$, which highlights the potential relevance of the proposed methods in practice. Given the identification result, general functionals of parameters of interest are also identified, since they can be written as functions of $q_t(\tau)$, $q_{At}(\tau)$, $q_{Bt}(\tau)$, and consequently as functions of the observable variables (Y, T, \mathbf{X}) . For example, the quantile treatment effect (QTE) will be $\Delta(\tau) = q_1(\tau) - q_0(\tau) = [1 \quad -1] Q(\tau, \tau)$ and for quantile treatment effect on the treated (QTT) will be $\Delta_A(\tau) = q_{A1}(\tau) - q_{A0}(\tau) = [1 \quad -1] Q_A(\tau, \tau)$. Less common than the previous two treatment effect parameters, the QTU, the quantile treatment effect on the untreated, is defined as $\Delta_B(\tau) = [1 \quad -1] Q_B(\tau, \tau) = q_{B1}(\tau) - q_{B0}(\tau)$. Other functionals, such as the Makarov bounds for the CDF of $Y(1) - Y(0)$ (Fan and Park (2010)) that explicitly depend on $Q_A(\tau, \tau')$ and $Q_B(\tau, \tau')$ at different points (τ, τ') , can similarly be obtained from*

the quantiles of potential outcomes.

3.2 Estimation

We are interested in estimation and inference for the $GIC(\tau)$ and $GIC^*(\tau)$. Equations (1) and (2) show that GIC can be written as a function of the quantiles of potential outcomes. Thus, we estimate each component of the vectors $Q(\tau, \tau')$, $Q_A(\tau, \tau')$ and $Q_B(\tau, \tau')$ to construct estimators for the $GIC(\tau)$ and $GIC^*(\tau)$.

Given identification, we are able to estimate the parameters of interest using a multi-step estimator as follows.

Step 1 Estimate $p(\mathbf{X})$ parametrically or nonparametrically and obtain an estimator $\hat{p}(\mathbf{X})$.⁵

The estimator of p is the sample average of T , i.e., $\hat{p} = n^{-1} \sum_{i=1}^n T_i$.

Step 2 For each $(\tau, \tau') \in \mathcal{T} \times \mathcal{T}$, obtain

$$\hat{Q}(\tau, \tau') = \begin{bmatrix} \hat{q}_1(\tau) \\ \hat{q}_0(\tau') \end{bmatrix}, \hat{Q}_A(\tau, \tau') := \begin{bmatrix} \hat{q}_{A1}(\tau) \\ \hat{q}_{A0}(\tau') \end{bmatrix}, \text{ and } \hat{Q}_B(\tau, \tau') := \begin{bmatrix} \hat{q}_{B1}(\tau) \\ \hat{q}_{B0}(\tau') \end{bmatrix},$$

where, for $t = 0, 1$, $\hat{q}_t(\tau)$, $\hat{q}_{At}(\tau)$ and $\hat{q}_{Bt}(\tau)$ satisfying the following conditions:

$$\mathbb{E}[\hat{w}_t(\tau - \mathbf{1}\{Y < \hat{q}_t(\tau)\})] = 0 \quad (3)$$

$$\mathbb{E}[\hat{w}_{At}(\tau - \mathbf{1}\{Y < \hat{q}_{At}(\tau)\})] = 0 \quad (4)$$

$$\mathbb{E}[\hat{w}_{Bt}(\tau - \mathbf{1}\{Y < \hat{q}_{Bt}(\tau)\})] = 0, \quad (5)$$

where $\hat{w}_{1,i} = T_i/\hat{p}(\mathbf{X}_i)$, $\hat{w}_{0,i} = (1 - T_i)/(1 - \hat{p}(\mathbf{X}_i))$, $\hat{w}_{A1,i} = T_i/\hat{p}$,

$\hat{w}_{A0,i} = [(1 - T_i)/(1 - \hat{p}(\mathbf{X}_i))] [\hat{p}(\mathbf{X}_i)/\hat{p}]$, $\hat{w}_{B1,i} = [T_i/\hat{p}(\mathbf{X}_i)] [(1 - \hat{p}(\mathbf{X}_i))/(1 - \hat{p})]$

and $\hat{w}_{B0,i} = (1 - T_i)/(1 - \hat{p})$.

In practice, estimators of $q_t(\tau)$, $q_{At}(\tau)$ and $q_{Bt}(\tau)$ can be obtained by weighted quantile

⁵Appendix 6.3 discusses the practical estimation of $p(\mathbf{X})$.

regressions (QR)

$$\widehat{q}_t(\tau) = \arg \min_q \mathbb{E} [\widehat{w}_{t,i} \rho_\tau (Y_i - q)], \quad (6)$$

$$\widehat{q}_{At}(\tau) = \arg \min_q \mathbb{E} [\widehat{w}_{At,i} \rho_\tau (Y_i - q)] \quad \text{and} \quad (7)$$

$$\widehat{q}_{Bt}(\tau) = \arg \min_q \mathbb{E} [\widehat{w}_{Bt,i} \rho_\tau (Y_i - q)], \quad (8)$$

where $\rho_\tau(u) := u(\tau - \mathbf{1}\{u < 0\})$ is the check function as in Koenker and Bassett (1978).

Step 3 Finally, we can plug-in estimates of the quantiles of the potential outcomes into the expressions to estimate GIC in (1) as following

$$\widehat{GIC}(\tau) = \frac{\widehat{q}_{A1}(\tau) - \widehat{q}_{B0}(\tau)}{\widehat{q}_{B0}(\tau)} = \frac{\begin{bmatrix} 1 & 0 \end{bmatrix} \widehat{Q}_A(\tau, \tau)}{\begin{bmatrix} 0 & 1 \end{bmatrix} \widehat{Q}_B(\tau, \tau)} - 1,$$

where we estimate $\widehat{q}_{A1}(\tau)$ and $\widehat{q}_{B0}(\tau)$ as in (7) and (8), respectively. To compute the corresponding weights, we estimate the propensity score, $p(\mathbf{X})$, by approximating its log-odds ratio by a polynomial and use the logistic link function with covariates given below in the data description.

Analogously, we can also estimate the counterfactual GIC^* in (2) as

$$\widehat{GIC}^*(\tau) = \frac{\widehat{q}_{B1}(\tau) - \widehat{q}_{B0}(\tau)}{\widehat{q}_{B0}(\tau)} = \frac{\begin{bmatrix} 1 & 0 \end{bmatrix} \widehat{Q}_B(\tau, \tau)}{\begin{bmatrix} 0 & 1 \end{bmatrix} \widehat{Q}_B(\tau, \tau)} - 1,$$

which, as described previously, is the growth incidence curve for quantile τ if the distribution of explanatory variables of income had remained fixed from period 0 to 1.

There are other alternative estimators available in the literature for the quantile objects of interest defined in Step 2 above. Donald and Hsu (2014) discuss an estimator that makes use of the inverse of the cumulative distribution function (CDF) of the potential outcomes. Their approach to estimate the quantiles is a three-step procedure. In the first step one needs to compute weights; in the second step the CDF is computed for all points on its support by using an inverse probability weighted estimator; and in the third step one obtains the quantile by inverting the CDF. We show below that the estimator proposed by Donald and Hsu (2014) and our proposed method are asymptotically equivalent. Nevertheless, the

estimator discussed in this paper has several practical advantages. First, our estimator for the quantiles is a two-step method: the first step coincides with the one of Donald and Hsu (2014), but unlike that method, our QR estimator for the object of interest is obtained without having to invert the CDF. This is possible because of the second advantage of our method: QR has a linear program representation, which makes practical computation simple and allows using weights directly into the objective function that is solved. Finally, if one is interested in quantiles, and its transformations, using the proposed estimator is attractive due to its computational efficiency and accuracy in finite samples.⁶

Remark 2. *One can also easily use the multi-step estimator defined above to obtain estimates for other functionals of interest. For example, the estimator of QTE will be $\widehat{\Delta}(\tau) = \widehat{q}_1(\tau) - \widehat{q}_0(\tau) = [1 \quad -1] \widehat{Q}(\tau, \tau)$ and for QTT will be $\widehat{\Delta}_A(\tau) = \widehat{q}_{A1}(\tau) - \widehat{q}_{A0}(\tau) = [1 \quad -1] \widehat{Q}_A(\tau, \tau)$. Other functionals, such as the Makarov bounds for the quantiles of the distribution of treatment effects, $Y(1) - Y(0)$, are estimated using the analytical expressions of these estimated bounds as functions of $\widehat{Q}_A(\tau, \tau')$ and $\widehat{Q}_B(\tau, \tau')$.*

3.3 Asymptotic properties

In this section, we derive the asymptotic properties of the multi-step estimator for the quantile process. We first focus on the properties of the estimator of $q_t(\tau)$ and establish the uniform consistency and the weak limit of $\widehat{q}_t(\tau)$, in $\ell^\infty(\mathcal{T})$. The extension to $\widehat{q}_{At}(\tau)$ and $\widehat{q}_{Bt}(\tau)$ is direct. We also establish the consistency and the weak limit of $\widehat{Q}(\tau, \tau')$, $\widehat{Q}_A(\tau, \tau')$ and $\widehat{Q}_B(\tau, \tau')$ in $\ell^\infty(\mathcal{T}) \times \ell^\infty(\mathcal{T})$. The asymptotic properties of the $\widehat{GIC}(\tau)$ and $\widehat{GIC}^*(\tau)$ follow from these results. In addition, we derive the uniform semiparametric efficiency of the estimator. Finally, we discuss how in practice we estimate weights used to compute $\widehat{q}_t(\tau)$. The two last results are collected in the Appendix.⁷

3.3.1 Consistency

Consistency is a desired property for most estimators. For the consistency of process $\widehat{q}_t(\tau)$ over $\tau \in \mathcal{T}$, consider the following conditions.

⁶We refer the reader to Koenker, Leorato, and Peracchi (2013) for a discussion and comparison on the statistical properties of the distribution regression and the quantile regression approaches.

⁷In Appendix 6.2, we provide results for the uniform semiparametric efficiency of the estimator. In Appendix 6.3 we discuss the practical estimation of the corresponding nuisance parameters, $w_t(\cdot)$, $w_{At}(\cdot)$, and $w_{Bt}(\cdot)$.

QC.I For $s, t \in \{0, 1\}$, the densities $f_{Y(s)|T}(\cdot|t)$ are bounded above and, uniformly in τ , positive. Also, for any $\delta > 0$,

$$\inf_{|q_t(\tau)|_\infty > \delta} |\mathbb{E}[w_t(T, \mathbf{X})(\tau - 1\{Y < q_t(\tau)\})]|_\infty > \epsilon_\delta.$$

QC.II There exists $0 < M_w < \infty$ such that $w_t(T, \mathbf{X}) < M_w$, a.e. (T, \mathbf{X}) .

QC.III $|\hat{w}_t - w_t|_\infty = o_p(1)$.

These conditions are standard in the literature. We state **QC.I** and **QC.II** for self-containedness. As usual in the QR literature, **QC.I** requires the density to be bounded away from infinity. The second part of **QC.I** is a standard identification condition. It is similar to Angrist, Chernozhukov, and Fernandez-Val (2006) and Firpo (2007), and it follows from **I.I–I.III** for each τ . **QC.II** imposes boundedness on the density of \mathbf{X} . It is analogue to Assumption 1(ii) of Firpo (2007) and follows directly from **I.III**. **QC.III** requires consistent estimation of the nuisance parameter. This is a usual requirement corresponding to (1.4) of Theorem 1 of Chen, Linton, and Van Keilegom (2003).

The following result establishes consistency of the estimator over the set of quantiles.

Theorem 1. *Suppose that $\mathbb{E}[w_t(T, \mathbf{X})m(Y, q_t(\tau); \tau)] = 0$, and that conditions **QC.I–QC.III** are satisfied. Then, for $t = 0, 1$, as $n \rightarrow \infty$*

$$\sup_{\tau \in \mathcal{T}} |\hat{q}_t(\tau) - q_t(\tau)| = o_{p^*}(1).$$

The extension of Theorem 1 to $\hat{q}_{At}(\cdot)$ and $\hat{q}_{Bt}(\cdot)$, $t = 0, 1$ is direct. The assumptions **QC.I–QC.III** are analogous.

3.3.2 Weak convergence

Now we derive the limiting distribution of the general $\hat{q}_t(\tau)$ estimator. We impose the following sufficient conditions.

QG.I The functions $\hat{w}_t(T, \mathbf{X}) \in \Pi$ and $\hat{w}_t(T, \mathbf{X}) \xrightarrow{p} w_t(T, \mathbf{X})$ uniformly in (T, \mathbf{X}) over compact sets, where $w_t(T, \mathbf{X}) \in \Pi$, and Π is a function class of uniformly smooth functions in (T, \mathbf{X}) with domain $\{0, 1\} \times \mathcal{X}$.

QG.II $\sqrt{n} (\mathbb{E}[(\widehat{w}_t(T, \mathbf{X}) - w_t(T, \mathbf{X}))(\tau - \mathbf{1}\{Y < q_t(\tau)\})] + \mathbb{E}[w_t(T, \mathbf{X})(\tau - \mathbf{1}\{Y < q_t(\tau)\})])$ converges weakly.

QG.III $|\widehat{w}_t(T, \mathbf{X}) - w_t(T, \mathbf{X})|_\infty = o_p(n^{-1/4})$.

Assumptions **QG.I–QG.III** concern the properties of the weights. They are high level conditions and will be discussed in the section of the estimation of w_t . Conditions **QG.I** and **QG.II** allow for estimated weights. Assumption **QG.II** is similar to Cattaneo (2010). Examples satisfying **QG.II** include smooth function classes. These assumptions allow for a wide variety of nonparametric and parametric estimators. **QG.III** strengthens **QC.III** such that the estimator of the nuisance parameter converges at a rate faster than $n^{-1/4}$. A similar assumption appears in Chen, Linton, and Van Keilegom (2003).

Now we present the weak convergence result.

Theorem 2. *For $t = 0, 1$, suppose that $|\mathbb{E}[w_t(T, \mathbf{X})m(Y, q_t(\tau); \tau)]|_\infty = 0$, that $|\widehat{q}_t - q_t|_\infty = o_p^*(1)$, and that conditions **QC.I–QC.II**, **QG.I–QG.III** are satisfied. Then, in $\ell^\infty(\mathcal{T})$,*

$$\sqrt{n}(\widehat{q}_t - q_t) \rightsquigarrow \mathbb{G}_t,$$

where \mathbb{G}_t is a mean zero Gaussian process with covariance function $\mathbb{E}[\mathbb{G}_t(\tau)\mathbb{G}_t(\tau')^\top] = D_t^{-1}(\tau)S_{tt}(\tau, \tau')[D_t^{-1}(\tau')]^\top$, with, for $t = 0, 1$, and $l = 0, 1$,

$$\begin{aligned} D_t(\tau) &= \frac{\partial \mathbb{E}[w_t(T, \mathbf{X})m(Y, q_t(\tau); \tau)]}{\partial q_t(\tau)} \Big|_{q_t(\tau)=q_t(\tau)} \\ S_{tt}(\tau, \tau') &= \mathbb{E}[(w_t(T, \mathbf{X})(m(Y, q_t(\tau); \tau) - \mathbb{E}[m(Y, q_t(\tau); \tau)|X, T = t]) + \mathbb{E}[m(Y, q_t(\tau); \tau)|X, T = t]) \\ &\quad \cdot (w_l(T, \mathbf{X})(m(Y, q_l(\tau'); \tau') - \mathbb{E}[m(Y, q_l(\tau'); \tau')|X, T = l]) + \mathbb{E}[m(Y, q_l(\tau'); \tau')|X, T = l])]. \end{aligned}$$

The result in Theorem 2 shows that the limiting distribution of the estimator is a Gaussian process. Thus, if one fixes a quantile at $\bar{\tau}$, then the limiting distribution collapses to a simple normal distribution, as in Firpo (2007). For practical inference, below we provide inference methods over the set of quantiles that are simple to implement in applications.⁸ As before, the extension of Theorem 2 to $\widehat{q}_{At}(\cdot)$ and $\widehat{q}_{Bt}(\cdot)$, $t = 0, 1$ is direct. The assumptions corresponding to **QG.I–QG.III** are analogous.

⁸Firpo and Pinto (2015) present a similar result to Theorem 2. Nevertheless our proof technique is different on the treatment of both infinite dimension parameters. In addition, we do not require compactness of the support of \mathbf{X} and impose weaker assumptions on \widehat{w}_t .

Given the result in Theorem 2, it is simple to establish the weak convergence to the vector $\widehat{Q}(\tau, \tau')$. The results for $\widehat{Q}_A(\tau, \tau')$ and $\widehat{Q}_B(\tau, \tau')$ are analogous.

Corollary 1. *Assume the conditions of Theorem 2, as $n \rightarrow \infty$, in $\ell^\infty(\mathcal{T}) \times \ell^\infty(\mathcal{T})$*

$$\sqrt{n}(\widehat{Q} - Q) = \sqrt{n} \begin{bmatrix} (\widehat{q}_1 - q_1) \\ (\widehat{q}_0 - q_0) \end{bmatrix} \rightsquigarrow \mathbb{G} = \begin{bmatrix} \mathbb{G}_1 \\ \mathbb{G}_0 \end{bmatrix},$$

where \mathbb{G} is the vector of Gaussian processes with covariance function

$$\mathbb{E}[\mathbb{G}(\tau, \tau')\mathbb{G}(\tau'', \tau''')^\top] = \begin{bmatrix} D_1^{-1}(\tau)S_{11}(\tau, \tau'')[D_1^{-1}(\tau'')]^\top & D_1^{-1}(\tau)S_{10}(\tau, \tau''')[D_0^{-1}(\tau''')]^\top \\ D_0^{-1}(\tau')S_{01}(\tau', \tau'')[D_1^{-1}(\tau'')]^\top & D_0^{-1}(\tau')S_{00}(\tau', \tau''')[D_0^{-1}(\tau''')]^\top \end{bmatrix}.$$

In order to perform inference on functions of the $Q(\tau, \tau')$, we impose a differentiability condition on such functions and state a functional delta method result. Consider the following assumption.

QG.IV (Hadamard) The functional $h : \ell^\infty(\mathcal{T}) \times \ell^\infty(\mathcal{T}) \rightarrow \ell^\infty(\mathcal{T})$ defined over the distribution of potential outcomes is Hadamard differentiable at Q , with Hadamard derivative given by $h(\cdot)'$.

The following result is a well known application of the functional delta method, we include it for completeness.

Lemma 1. *Assume the conditions of Theorem 2, and **QG.IV**, as $n \rightarrow \infty$,*

$$\sqrt{n}(h(\widehat{Q}) - h(Q)) \rightsquigarrow h(\mathbb{G})'.$$

Donald and Hsu (2014) establish the weak convergence of a quantile estimator that makes use of the inverse of the CDF in their Theorem 3.8. Their result is similar to that in Theorem 2 above. Nevertheless, as mentioned previously, the quantile estimators are different. In addition, the assumptions required to establish the results are different. On the one hand, Donald and Hsu (2014) impose strong conditions to derive the result. For instance, their Assumption 3.1 requires that the distributions of $Y(0)$ and $Y(1)$ have convex and compact supports. Their Assumption 3.2 requires all the covariates to be continuous

and the support of the vector of covariates, \mathbf{X} , to be compact. We are able to somewhat relax these assumptions. Given that we work with a standard semiparametric estimator, and a quantile regression framework in the second step, we do not require such assumptions to derive the asymptotic properties of our proposed estimator.

Now we to return to the main object of interest and analyze the growth incidence curves, $GIC(\tau)$ and $GIC^*(\tau)$. As an application of Theorem 2 and Lemma 1, we derive the asymptotic distribution for $GIC(\tau)$. Corollary 1 implies that

$$\sqrt{n}(\widehat{Q}_A - Q_A) \rightsquigarrow \mathbb{G}_A, \quad \text{and} \quad \sqrt{n}(\widehat{Q}_B - Q_B) \rightsquigarrow \mathbb{G}_B,$$

where $\mathbb{G}_A(\tau)$ and $\mathbb{G}_B(\tau)$ are Gaussian processes with variance-covariance functions that can be obtained as an application of Corollary 1.

Recall that $GIC(\tau) = \frac{[1 \ 0]Q_A(\tau, \tau)}{[0 \ 1]Q_B(\tau, \tau)} - 1$, and $GIC^*(\tau) = \frac{[1 \ 0]Q_B(\tau, \tau)}{[0 \ 1]Q_B(\tau, \tau)} - 1$. These functionals are differentiable at (Q_A, Q_B) , as long as $q_{B0}(\tau) \neq 0$ with derivatives defined by

$$GIC(\mathbb{G}_A, \mathbb{G}_B)' = \frac{1}{[0 \ 1]Q_B} [1 \ 0] \mathbb{G}_A - \frac{[1 \ 0]Q_A}{([0 \ 1]Q_B)^2} [0 \ 1] \mathbb{G}_B,$$

and for GIC^* we have that

$$GIC^*(\mathbb{G}_B)' = \left(\frac{1}{[0 \ 1]Q_B} [1 \ 0] - \frac{[1 \ 0]Q_B}{([0 \ 1]Q_B)^2} [0 \ 1] \right) \mathbb{G}_B.$$

Therefore, from a functional delta method we have the following results.

Corollary 2. *Assume the conditions of Theorem 2, as $n \rightarrow \infty$, in $\ell^\infty(\mathcal{T})$*

$$\sqrt{n}(\widehat{GIC} - GIC) \rightsquigarrow GIC(\mathbb{G}_A, \mathbb{G}_B)' \tag{9}$$

$$\sqrt{n}(\widehat{GIC}^* - GIC^*) \rightsquigarrow GIC^*(\mathbb{G}_B)'. \tag{10}$$

3.4 Inference procedures

In this section, we turn our attention to inference procedures on the GIC . Important questions posed in the econometric and statistical literatures concern the nature of the

impact of a policy intervention or treatment on the outcome distributions of interest. The corresponding questions for the GIC are, for example, whether there is significant income growth at any quantile (the null hypothesis being $GIC(\tau) = 0$ for all τ); and whether growth is uniform or heterogeneous ($GIC(\tau)$ equals the average growth rate, for all τ). One can also ask if growth is non-decreasing in τ ($GIC(\tau)' \geq 0$ for all τ). Since the main objective of this paper is to study the growth incidence curve, and these questions and hypotheses are formulated for the entire GIC process, we develop inference procedures for the quantile process over the set of quantiles indexed by τ .

3.4.1 Test statistics

We seek to develop inference for GIC over the index set of quantiles \mathcal{T} . We present results for functionals of quantiles of the marginal distributions of potential outcomes, and in particular, the $GIC(\tau)$ and $GIC^*(\tau)$. Let $\beta(\tau)$ be a functional of Q , Q_A , and Q_B , that is, $\beta(\tau) = h(Q(\tau, \tau))$. In particular, we are interested in $\beta(\tau) = GIC(\tau) = \frac{[1 \ 0]Q_A(\tau, \tau)}{[0 \ 1]Q_B(\tau, \tau)} - 1$, and the counterfactual one $\beta(\tau) = GIC^*(\tau) = \frac{[1 \ 0]Q_B(\tau, \tau)}{[0 \ 1]Q_B(\tau, \tau)} - 1$.

We discuss three main hypotheses of interest. First, we consider the following standard null hypothesis

$$H_0 : \beta(\tau) - r(\tau) = 0, \quad \tau \in \mathcal{T}, \quad (11)$$

uniformly, where the vector $r(\tau)$ is assumed to be known, continuous in τ over \mathcal{T} , and $r \in \ell^\infty(\mathcal{T})$. More generally, the hypothesis in (11) embeds several interesting hypotheses about the parameters of the quantile function.

Example (The uniformly null effect hypothesis). *A basic hypothesis is that the growth incidence curve, $GIC(\tau)$, is statistically equal to zero for all $\tau \in \mathcal{T}$. The alternative is that the it differs from zero at least for some $\tau \in \mathcal{T}$. In this case, $r(\tau) = 0$, and relative inequality remains stable.*

The basic inference process to test the null hypothesis (11) is

$$W_n(\tau) := \widehat{\beta}(\tau) - r(\tau), \quad \tau \in \mathcal{T}.$$

To derive the asymptotic properties of the above statistic, we need to compute the estimator $\widehat{\beta}(\tau)$, which is given by $\widehat{\beta} = h(\widehat{Q})$. The $GIC(\tau)$ estimate is $\widehat{\beta}(\tau) = \widehat{GIC}(\tau) = \frac{[1 \ 0]\widehat{Q}_A(\tau, \tau)}{[0 \ 1]\widehat{Q}_B(\tau, \tau)} - 1$, and the estimate for $GIC^*(\tau)$ is $\widehat{\beta}(\tau) = \widehat{GIC}^*(\tau) = \frac{[1 \ 0]\widehat{Q}_B(\tau, \tau)}{[0 \ 1]\widehat{Q}_B(\tau, \tau)} - 1$, which for a fixed quantile τ , has an asymptotic normal distribution as given in Corollary 2.

General hypotheses about $\beta(\tau)$ can be accommodated through functions of $W_n(\cdot)$. We consider the Kolmogorov-Smirnov and Cramér-von Mises type test statistics, $V_n = f(W_n(\cdot))$, where $f(\cdot)$ is a general functional of the process $W_n(\cdot)$. In particular, we consider different functionals that lead to different test statistics, such as

$$V_{1n} := \sqrt{n} \sup_{\tau \in \mathcal{T}} |W_n(\tau)|, \quad V_{2n} := \sqrt{n} \int_{\tau \in \mathcal{T}} |W_n(\tau)| d\tau.$$

There are many alternative possible statistics as: $V_{3n} := \sqrt{n} \sup_{\tau \in \mathcal{T}} W_n(\tau)^2$ and $V_{4n} := \sqrt{n} \int_{\tau \in \mathcal{T}} W_n(\tau)^2, d\tau$, among others. In this paper we concentrate on V_{1n} and V_{2n} .

These statistics and their associated limiting theory provide a natural foundation for testing the null hypothesis. Now we present the limiting distributions of the test statistics under the null hypothesis. From Corollary 1 and Lemma 1 under the null hypothesis ($H_0 : \beta = h(Q) = r$), it follows that $\sqrt{n}(h(\widehat{Q}) - h(Q)) \rightsquigarrow h(\mathbb{G})'$. Thus, the following lemma summarizes the limiting distributions.

Lemma 2. *Assume the conditions of Theorem 2, and **QG.IV**. Under $H_0 : \beta(\tau) = h(Q(\tau)) = r(\tau)$, $\tau \in \mathcal{T}$, as $n \rightarrow \infty$,*

$$V_{1n} \rightsquigarrow \sup_{\tau \in \mathcal{T}} |h(\mathbb{G}(\tau))'|, \quad V_{2n} \rightsquigarrow \int_{\tau \in \mathcal{T}} |h(\mathbb{G}(\tau))'| d\tau.$$

When performing tests for the *GIC*, the limiting distributions of the test statistics under the null hypothesis follows from Theorem 2. Under the null hypothesis ($H_0 : GIC(\tau) = r(\tau)$), it follows $\sqrt{n}(\widehat{GIC}(\tau) - r(\tau)) \rightsquigarrow GIC(\mathbb{G}_A, \mathbb{G}_B)'$. Thus, the following corollary summarizes the limiting distributions. The result for $H_0 : GIC^*(\tau) = r(\tau)$ is analogous.

Corollary 3. *Assume the conditions of Theorem 2. Under $H_0 : GIC(\tau) = r(\tau)$, as $n \rightarrow \infty$,*

$$V_{1n} \rightsquigarrow \sup_{\tau \in \mathcal{T}} |GIC(\mathbb{G}_A, \mathbb{G}_B)'|, \quad V_{2n} \rightsquigarrow \int_{\tau \in \mathcal{T}} |GIC(\mathbb{G}_A, \mathbb{G}_B)'| d\tau.$$

The second hypothesis of interest concerns an unknown $r(\tau)$, which needs to be estimated. In many examples of interest, the component $r(\tau)$ in the null hypothesis (11) is unknown or defined as a function of the conditional distribution and thus needs to be estimated (see, e.g., Koenker and Xiao (2002) and Chernozhukov and Fernandez-Val (2005)). $r(\tau)$ might,

for example, be $GIC(\tau)$ for a different country, or period. Or it might be $GIC^*(\tau)$. The natural expedient of replacing the unknown r in the test statistic by estimates introduces some fundamental difficulties. The estimate will be denoted as $\hat{r}(\tau)$. Let

$$\bar{W}_n(t) := \hat{\beta}(\tau) - \hat{r}(\tau), \quad \tau \in \mathcal{T}.$$

In this framework, we follow Chernozhukov and Fernandez-Val (2005) and assume that the quantile estimates and nuisance parameter estimates satisfy the following: \sqrt{n} -consistent estimators for $\hat{\beta}(\cdot)$ and $\hat{r}(\cdot)$, such that $\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot)) \rightsquigarrow h(\mathbb{G}(\cdot))'$ and $\sqrt{n}(\hat{r}(\cdot) - r(\cdot)) \rightsquigarrow \mathbb{G}_r(\cdot)$ jointly in $\ell^\infty(\mathcal{T})$, where $(h(\mathbb{G}(\cdot))', \mathbb{G}_r(\cdot))$ is a zero mean continuous Gaussian process with a non-degenerate covariance kernel. Thus, we have that $\sqrt{n}(\hat{\beta}(\tau) - \hat{r}(\tau)) \rightsquigarrow h(\mathbb{G}(\tau))' - \mathbb{G}_r(\tau)$. The process remains asymptotically Gaussian; however, the estimation of $r(\tau)$ introduces a new drift component that additionally complicates the covariance kernel of the process.

Under the null hypothesis $H_0 : \beta(\tau) = r(\tau)$, the test statistics become:

$$\bar{V}_{1n} := \sqrt{n} \sup_{\tau \in \mathcal{T}} |\bar{W}_n(\tau)|, \quad \bar{V}_{2n} := \sqrt{n} \int_{\tau \in \mathcal{T}} |\bar{W}_n(\tau)| d\tau.$$

Example (The uniformly constant (but unknown) effect hypothesis). *A basic hypothesis is that the growth incidence curve, $GIC(\tau)$, is statistically equal to the mean growth rate for all $\tau \in \mathcal{T}$, i.e., growth has no distributional heterogeneity. The alternative is that $GIC(\tau)$ differs from the mean at least for some $\tau \in \mathcal{T}$. In this case, $\hat{r}(\tau) = \hat{\gamma}_{AGR}$, (where γ_{AGR} is the mean growth rate).*

Now we display the limiting distributions of these test statistics under the null hypothesis.

Lemma 3. *Assume the conditions of Theorem 2 and that $\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot)) \rightsquigarrow h(\mathbb{G}(\cdot))'$ and $\sqrt{n}(\hat{r}(\cdot) - r(\cdot)) \rightsquigarrow \mathbb{G}_r(\cdot)$ jointly in $\ell^\infty(\mathcal{T})$, where $(h(\mathbb{G}(\cdot))', \mathbb{G}_r(\cdot))$ is a zero mean continuous Gaussian process with a non-degenerate covariance kernel.. Under $H_0 : \beta(\tau) = h(Q(\tau)) = r(\tau)$, $\tau \in \mathcal{T}$, as $n \rightarrow \infty$,*

$$\bar{V}_{1n} \rightsquigarrow \sup_{\tau \in \mathcal{T}} |h(\mathbb{G}(\tau))' - \mathbb{G}_r(\tau)|, \quad \bar{V}_{2n} \rightsquigarrow \int_{\tau \in \mathcal{T}} |h(\mathbb{G}(\tau))' - \mathbb{G}_r(\tau)| d\tau.$$

This result can be applied to test for the GIC . The limiting distributions of the test statistics under the null hypothesis follow from Lemma 3. Under the null hypothesis ($H_0 :$

$GIC(\tau) = r(\tau)$), it follows $\sqrt{n}(\widehat{GIC}(\tau) - r(\tau)) \rightsquigarrow GIC(\mathbb{G}_A, \mathbb{G}_B)'$, and $\sqrt{n}(\widehat{r}(\tau) - r(\tau)) \rightsquigarrow \mathbb{G}_r(\tau)$. The following corollary summarizes the limiting distributions. The result for $H_0 : GIC^*(\tau) = r(\tau)$ is analogous.

Corollary 4. *Assume the conditions of Lemma 3, with $\beta(\tau) = GIC(\tau)$, as $n \rightarrow \infty$,*

$$\bar{V}_{1n} \rightsquigarrow \sup_{\tau \in \mathcal{T}} |GIC(\mathbb{G}_A, \mathbb{G}_B)' - \mathbb{G}_r(\tau)|, \quad \bar{V}_{2n} \rightsquigarrow \int_{\tau \in \mathcal{T}} |GIC(\mathbb{G}_A, \mathbb{G}_B)' - \mathbb{G}_r(\tau)| d\tau.$$

Finally, we consider testing hypotheses concerning inequalities on both null and alternative hypotheses as

$$H_0 : \beta(\tau) \geq 0 \quad vs \quad H_1 : \beta(\tau) < 0, \quad \tau \in \mathcal{T}. \quad (12)$$

The following is an example of hypotheses that may be considered.

Example (The first-order stochastic dominance hypothesis). *An important practical hypothesis involves the composite null $GIC(\tau) \geq r(\tau)$, for all $\tau \in \mathcal{T}$, versus the alternative of $GIC(\tau) < r(\tau)$, for some $\tau \in \mathcal{T}$. When $r(\tau) = 0$ and because $GIC(\tau) = \frac{q_{A1}(\tau) - q_{B0}(\tau)}{q_{B0}(\tau)}$, such that for $q_{B0}(\tau) \neq 0$, testing whether $GIC(\tau) \geq 0$ is equivalent to test whether $q_{A1}(\tau) \geq q_{B0}(\tau)$, ie, that $F_{Y(1)|T=1}$ stochastically dominates $F_{Y(0)|T=0}$ in first-order.*

Therefore, the above example describes a test which is analogous to a first order stochastic dominance as in Donald and Hsu (2014). These null hypotheses of interest can be formalized as $H_0 : \beta(\tau) \geq 0$, and the test statistic becomes:

$$\tilde{V}_{1n} := \sqrt{n} \sup_{\tau \in \mathcal{T}} \tilde{W}_n(\tau),$$

where $\tilde{W}_n(\tau) = \widehat{\beta}(\tau)$.

We employ the test statistic \tilde{V}_{1n} since it has been known in the literature that when the null hypothesis involves an inequality, the set of points satisfying the null hypothesis is usually not a singleton (see, e.g., Linton, Maasoumi, and Whang (2005)). The typical way to resolve this is to apply the least favorable configuration (LFC) to find a point in the null hypothesis least favorable to the alternative hypothesis. Hence, to derive the asymptotic properties of the above statistic, \tilde{V}_{1n} , one computes the estimator $\widehat{\beta}(\tau)$ and plugs it in, and given the LFC the limiting distribution is analogous to that in Lemma 2 and Corollary 2.

To perform practical inference we suggest the use of resampling techniques to approximate the limiting distributions and obtain critical values. To obtain the critical value for the first two criteria we use a bootstrap procedure, and for the inequality test we make use of subsampling.

3.4.2 Practical implementation of testing procedures

Implementation of the proposed tests in practice is simple. First, we discuss the test H_0 in (11). To implement the tests one needs to compute the statistics of test V_{1n} or V_{2n} . Analogously, when $r(\tau)$ is unknown, one computes \bar{V}_{1n} or \bar{V}_{2n} . We suggest the use of a recentered bootstrap procedure to calculate critical values. The steps for implementing the tests in practice are as follows.

First, the estimates of $\beta(\tau)$ are computed by solving the problems in equations (6)–(8) and calculating $\hat{\beta}(\tau)$. Second, W_n is calculated by centralizing $\hat{\beta}(\tau)$ at $r(\tau)$, and V_{1n} or V_{2n} is computed by taking the maximum over τ (V_{1n}) or summing over τ (V_{2n}). For the general case with unknown $r(\tau)$, the tests are computed in the same fashion. The only adjustment is the use of $\hat{r}(\tau)$ to compute \bar{W}_n . Third, after obtaining the test statistic, it is necessary to compute the critical values. We propose the following scheme. We use the test statistic V_{1n} as an example, but the procedure is the same for the other cases. Take B as a large integer. For each $b = 1, \dots, B$:

- (i) Obtain the resampled data $\{(Y_i^b, T_i^b, \mathbf{X}_i^b), i = 1, \dots, n\}$.
- (ii) Estimate $\hat{\beta}^b(\tau)$ and set $W_n^b(\tau) := (\hat{\beta}^b(\tau) - \hat{\beta}(\tau))$.
- (iii) Compute the test statistic of interest $\hat{V}_{1n}^b = \max_{\tau \in \mathcal{T}} \sqrt{n} |W_n^b(\tau)|$.

Let $\hat{c}_{1-\alpha}^B$ denote the empirical $(1 - \alpha)$ -quantile of the simulated sample $\{\hat{V}_{1n}^1, \dots, \hat{V}_{1n}^B\}$, where $\alpha \in (0, 1)$ is the nominal size. We reject the null hypothesis if V_{1n} is larger than $\hat{c}_{1-\alpha}^B$. In practice, the maximum in step (iii) is taken over a discretized subset of \mathcal{T} .

A formal justification the simulation method is stated as follows. Consider the following conditions.

QG.IB For any $\delta_n \downarrow 0$, $\sup_{\|w\|_{\Pi} \leq \delta_n} \left| \frac{1}{n} \sum_{i=1}^n w_t(T_i, \mathbf{X}_i) - \mathbb{E}[w_t(T, \mathbf{X})] \right|_{\infty} = o_{P^*}(1/\sqrt{n})$.

QG.IIB $\sqrt{n} \frac{1}{n} \sum_{i=1}^n [(\tau - \mathbf{1}\{Y_i < q_t(\tau)\})(\hat{w}_t^*(T_i, \mathbf{X}_i) - \hat{w}_t(T_i, \mathbf{X}_i))]$ converges weakly to a tight random element \mathbb{G} in $\ell^\infty(\mathcal{T})$ in P^* -probability.

Theorem 3. Under **QC.I–QC.II**, **QG.IB–QG.IIB** and **QG.III** with “in probability” replaced by “almost surely”, then, for $t = 0, 1$, the bootstrap estimator $\sqrt{n}(\widehat{q}_t^*(\tau) - \widehat{q}_t(\tau)) \rightsquigarrow \mathbb{G}(\tau)$ in P^* -probability in $\ell^\infty(\mathcal{T})$.

Theorem 3 establishes the consistency of the bootstrap procedure. It is important to highlight the connection between this result and the previous section. In fact, Theorem 3 shows that the limiting distribution of the bootstrap estimator is the same as that of Theorem 2, and hence the above resample scheme is able to mimic the asymptotic distribution of interest.

Now we move our attention to testing the H_0 displayed in (12). As discussed in Linton, Maasoumi, and Whang (2005), even when the data are i.i.d. the standard bootstrap might not work well when testing the inequality under the null hypothesis. This is because one needs to impose the null, which is difficult because it is defined by a complicated system of inequalities. Thus, we follow Linton, Maasoumi, and Whang (2005) and suggest the use of a subsampling method, which is very simple to define and yet provides consistent critical values.

We first define the subsampling procedure. Let $Z_i = \{(Y_i, T_i, \mathbf{X}_i) : i = 1, \dots, n\}$ and construct all possible subsets of size b . The number of such subsets B_n is “ n choose b .” Let S_n denote our test statistic \tilde{V}_{1n} computed over the entire sample. With some abuse of notation, the test statistic S_n can be re-written as a function of the data $\{Z_i : i = 1, \dots, n\}$:

$$S_n = \sqrt{n}s_n(Z_1, \dots, Z_n),$$

where $s_n(Z_1, \dots, Z_n)$ is given by $\sup_{\tau \in \mathcal{T}}[\tilde{W}_n(\tau)]$, where $\tilde{W}_n(\tau) = \widehat{\beta}(\tau)$. Let

$$J_n(w) = P(\sqrt{n}s_n(Z_1, \dots, Z_n) \leq z)$$

denote the distribution function of S_n . Let $s_{n,b,i}$ be equal to the statistic s_b evaluated at the subsamples $\{Z_i, \dots, Z_{i+b-1}\}$ of size b , i.e.

$$s_{n,b,i} = s_b(Z_i, Z_{i+1}, \dots, Z_{i+b-1}) \quad \text{for } i = 1, \dots, n - b + 1.$$

This means that we have to recompute $\widehat{q}_t(Z_i, Z_{i+1}, \dots, Z_{i+b-1})$ using each subsample as well. We note that each subsample of size b (taken without replacement from the original data)

is indeed a sample of size b from the true sampling distribution of the original data. Hence, it is clear that one can approximate the sampling distribution of S_n using the distribution of the values of $s_{n,b,i}$ computed over $n - b + 1$ different subsamples of size b . That is, we approximate the sampling distribution J_n of S_n by

$$\widehat{J}_{n,b}(w) = \frac{1}{n - b + 1} \sum_{i=1}^{n-b+1} 1 \left(\sqrt{b} s_{n,b,i} \leq w \right).$$

Let $g_{n,b}(1 - \alpha)$ denote the $(100 - \alpha)$ -th sample quantile of $\widehat{J}_{n,b}(\cdot)$. We call it the subsample critical value of significance level α . Thus, we reject the null hypothesis at the significance level α if $S_n > g_{n,b}(1 - \alpha)$. The computation of this critical value is not particularly onerous, although it depends on how big b is.

The validity of the subsampling methods for the quantile regression process was established by Chernozhukov and Fernandez-Val (2005).

A Supplemental Appendix collects Monte Carlo simulations conducted to evaluate the finite sample properties of the proposed tests. We conduct simulations to evaluate the performance of these tests in terms of size and power. The results provide evidence that the empirical levels of the tests approximate well the nominal or theoretical levels. Moreover, the tests possess large power against selected alternatives. The results are improved when the sample size increases, nevertheless they are not very sensitive to the numbers of bootstraps.

4 Wage distribution dynamics in the United States and Brazil, 1995-2007

This section illustrates the usefulness of the proposed methods with an empirical example. We compute the GIC and GIC^* for the two most populous nations in the Western Hemisphere, namely the United States and Brazil, for the 1995-2007 period, and compare results. In particular, we emphasize the role of the following decomposition of the GIC , introduced in Section 2 and reproduced below:

$$GIC(\tau) = GIC^*(\tau) + GIC^{**}(\tau) \cdot \frac{q_{B1}(\tau)}{q_{B0}(\tau)}.$$

The first term in this decomposition is the counterfactual GIC , which keeps the joint

distribution of observed covariates fixed (see equation 2). Under Assumptions **I.I–I.III**, this term can be interpreted as describing the growth process that would have obtained in the absence of any changes in that joint distribution. The second term of the decomposition is correspondingly interpreted as the effect of changes in the joint distribution of covariates.

Our reweighting method allows for the direct construction of the counterfactual GIC , with no need to postulate a structural relationship between wages, covariates and unobserved terms, as was required by the earlier literature that followed Juhn, Murphy, and Pierce (1993). Under that approach, economists would typically estimate OLS regressions for the two time periods separately and then construct a counterfactual wage distribution using estimated parameters and residuals from time $t = 1$ but covariates from time $t = 0$. This would yield a counterfactual distribution of wages at time $t = 1$, with a distribution of covariates that was fixed at time $t = 0$ (see, for example, Bourguignon, Ferreira, and Leite (2008)). In addition to requiring strong functional form assumptions, however, it is not clear how one would perform statistical inference on the counterfactual GIC using that method.

In this section we report the estimates for GIC and its counterfactual counterpart GIC^* , $\widehat{GIC}(\tau)$ and $\widehat{GIC}^*(\tau)$ respectively, over $\tau \in \mathcal{T}$. We also report the corresponding growth rates in average wages, γ_{AGR} and γ_{AGR^*} , respectively, for comparison. Moreover, using the techniques developed in the previous section, we perform inference on both sets of curves. Specifically, we apply the uniform tests, Kolmogorov-Smirnov (KS) and Cramér-von Mises (CVM), to test the following hypotheses:

- (i) Constant distribution: ($H_0 : GIC(\tau) = 0$ versus $H_A : GIC(\tau) \neq 0$);
- (ii) Distribution-neutral growth ($H_0 : GIC(\tau) = \gamma_{AGR}$ versus $H_A : GIC(\tau) \neq \gamma_{AGR}$, where γ_{AGR} is the growth rate in the average wage);
- (iii) Constant distribution, conditional on covariates, ($H_0 : GIC^*(\tau) = 0$ versus $H_A : GIC^*(\tau) \neq 0$);
- (iv) Distribution-neutral growth, conditional on covariates ($H_0 : GIC^*(\tau) = \gamma_{AGR^*}$ versus $H_A : GIC^*(\tau) \neq \gamma_{AGR^*}$, where γ_{AGR^*} is the counterfactual growth rate in average wage).

4.1 Data

4.1.1 CPS – US

Our data for the United States come from the March supplement to the Current Population Surveys (CPS) for 1995 and 2007.⁹ The dataset provides the distribution of labor earnings in the US in 1995 and 2007 for full-time workers of both genders. We use the following variables for our analysis. Y denotes real hourly labor earnings (sum of annual pretax wages, salaries, tips, and bonuses, divided by the number of hours worked annually). The vector \mathbf{X} consists of three covariates of Y , namely: (i) the worker’s age in years; (ii) a gender dummy; and (iii) a categorical variable for the highest educational level attained (“high school”, “some college” or “college”).

We restrict the sample to individuals aged 16 to 65 that reported a positive value for real hourly earnings in the previous year. Individuals with missing values for any of the four variables in Y and \mathbf{X} were excluded from the sample. After applying these filters, we trimmed the sample by dropping the top and bottom 0.5% of the distribution of hourly wages in each year, to eliminate outliers. Hourly wages are in US dollars of March 2007. The Consumer Price Index was used to inflate 1995 incomes: nominal values in 1995 were multiplied by 1.36 to be expressed in 2007 dollars. The final sample contains a total of 165,245 observations. Summary statistics are presented in Table 1.

Table 1: Summary Statistics – US

	Mean	S.D.	Min.	Max.	Observations
CPS 1995					
Hourly Work Earnings	16.730	11.402	1.047	74.162	69,494
Age	37.039	12.090	16	65	69,494
Male	0.527				69,494
High School	0.333				69,494
College Incomplete	0.295				69,494
College	0.243				69,494
CPS 2007					
Hourly Work Earnings	19.626	16.748	1.202	168.280	95,751
Age	39.284	12.807	16	65	95,751
Male	0.525				95,751
High School	0.305				95,751
College Incomplete	0.291				95,751
College	0.297				95,751

⁹<http://www.census.gov/programs-surveys/cps.html>

4.1.2 PNAD – Brazil

The Brazilian data come from the Pesquisa Nacional por Amostra de Domicílios (PNAD), an annual Brazilian household survey that samples households across (almost) the entire country.¹⁰ It collects information on various household characteristics, as well as individual incomes and education levels. We use PNAD data for 1995 and 2007. For comparability, we use the same four variables as for the CPS: real hourly labor earnings, age, gender, and a categorical educational attainment variable. IBGE, the Brazilian Statistics Bureau that is responsible for running PNAD, started including the rural Northern region in the PNAD sample after 2004 but, for comparability across years, we do not use information on the rural North for 2007.

As for the US, we restrict the sample to individuals aged 16 to 65 that reported a positive labor income in the previous year. Individuals with missing values for income or any of our three covariates were excluded from the sample. The top and bottom 0.5% of the distribution of hourly wages in each year was trimmed, as in the CPS. Hourly wages are in Brazilian reais (BRL) of September 2007, which means that nominal values in 1995 were multiplied by 2.89 to be expressed in 2007 prices. The final sample contains a total of 275,749 observations. Summary statistics are presented in Table 2.

Table 2: Summary Statistics – Brazil

	Mean	S.D.	Min.	Max.	Observations
PNAD 1995					
Hourly Work Earnings	5.565	7.312	0.300	61.317	119,770
Age	34.924	11.936	16	65	119,770
Male	0.631				119,770
High School	0.131				119,770
College Incomplete	0.046				119,770
College	0.064				119,770
PNAD 2007					
Hourly Work Earnings	5.659	6.978	0.312	62.500	155,979
Age	36.367	12.025	16	65	155,979
Male	0.589				155,979
High School	0.259				155,979
College Incomplete	0.085				155,979
College	0.094				155,979

A comparison of Tables 1 and 2 reveals considerable differences between the two labor

¹⁰In 1995, PNAD did not survey households in the rural areas of Acre, Amapá, Amazonas, Pará, Rondônia and Roraima — six states in the Amazon region.

forces. US full-time workers are on average some three years older than their Brazilian counterparts, and earn much higher wages: the nominal exchange rate in September 2007 was 1.90 BRL to the USD, so average wages in 2007 in this sample were approximately 6.6 times higher in the US than in Brazil. US workers are also much more educated, and the female share of the labor force is higher there. Over the twelve years between 1995 and 2007, both labor forces became a little older, and more educated. Educational attainment rose in both countries, but more markedly in Brazil, which started from a much lower level. Completion of high school in Brazil almost doubled over the period, and the college-educated share also rose from 6.4% to 9.4%. The female share of the labor force was essentially stable at 47% in the US, but rose from 37% to 41% in Brazil, driven primarily by a higher rate of female labor force participation (Ferreira, Firpo, and Messina (2016)).

4.2 Results

Before we present results for the *GIC*, we compute standard inequality measures for hourly real wages for both countries. Table 3 summarizes some of the main changes in the wage distributions in the US and Brazil over this period. The first panel presents five common measures of relative wage inequality for the two countries in 1995 and 2007, as well as for the counterfactual wage distribution $F_{Y(1)|T}(\cdot|0)$. The inequality measures are the Gini coefficient, the Theil-T index (that is, the Generalized Entropy measure with parameter = 1), the mean log deviation (also known as Theil-L, or $GE(0)$), the relative mean deviation, and the standard deviation of logarithms. The second panel presents the growth rate in mean hourly wages (γ_{AGR}) and the average of quantile-specific growth rates, across quantiles, denoted Mean *GIC*.

Table 3: Inequality measures hourly real wages (HRW) – US and Brazil

	US			Brazil		
	Factual	2007	Counter Factual	Factual	2007	Counter Factual
Gini	0.355	0.383	0.381	0.539	0.490	0.473
Theil Entropy	0.205	0.260	0.258	0.536	0.457	0.444
Theil mean log deviation	0.218	0.254	0.250	0.511	0.411	0.383
Relative mean deviation	0.257	0.274	0.272	0.404	0.366	0.348
Standard deviation of logs	0.682	0.710	0.703	0.963	0.852	0.814
Growth of mean wage (γ_{AGR})	0.173		0.101	0.017		-0.150
Mean <i>GIC</i>	0.127		0.063	0.138		-0.016

Below we discuss the findings for each country separately, including the KS and CVM tests of the four hypotheses listed earlier, before briefly comparing results across countries.

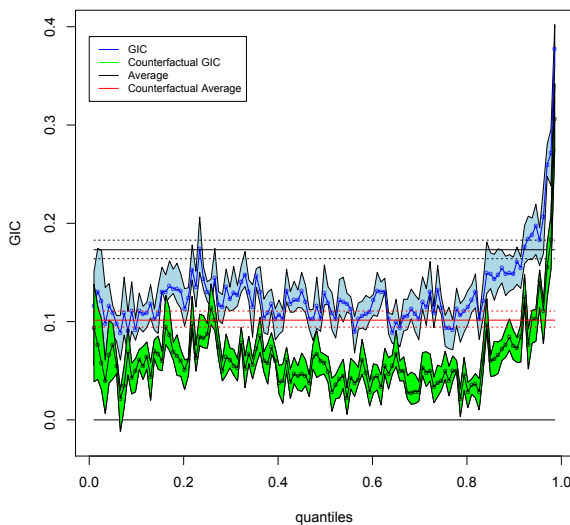
4.2.1 United States

Figure 1 presents the estimates for the GIC and GIC^* , and their corresponding 95% confidence intervals, for the US in the period 1995-2007. The blue line displays the GIC , and the straight horizontal black line represents the corresponding average growth rate, γ_{AGR} . The green line displays the counterfactual growth incidence curve, GIC^* , and the dashed red line shows its corresponding mean effect, γ_{AGR^*} . The number of bootstrap replications used to construct confidence intervals is 300.

The growth incidence curve for the US is essentially flat at a cumulative growth rate of approximately 10% for the first eight deciles of the distribution. From $\tau = 0.8$ onwards it begins to slope upwards, and the slope increases sharply for the uppermost decile. A growth rate of 10% over twelve years translates into an average annual wage growth rate of less than one percent over the period, supporting earlier descriptions of wage stagnation for most US workers, even during the “Goldilocks” economy that preceded the great financial crisis of 2008-09 (see, e.g. Kopczuk, Saez, and Song (2010) and Mishel, Bivens, Gould, and Shierholz (2012)). The fact that the growth in the average wage was considerably higher, at 17.3%, reflects the much better performance of the top quintile. This is also why it was higher than the average quantile-specific growth rate across quantiles, of 12.7%. The more rapid growth of wages among the top fifth of full-time workers naturally translated into rising inequality, as shown by all five inequality measures in Table 3. The commonly used Gini coefficient rose by almost three percentage points.

The basic finding that there was positive but heterogeneous wage growth in the US is found to be statistically significant by the inference results for the formal hypotheses formulated earlier, namely constant distribution and distribution-neutral growth for the GIC . These results are presented in Table 4, which reports the Kolmogorov-Smirnov (KS) and Cramér-von Mises tests (CVM) (V_{1n} and V_{2n} , respectively). First, we test the constant distribution hypothesis for the GIC uniformly over quantiles ($H_0 : GIC(\tau) = 0$), which is rejected at the 1% level of significance for both tests. Thus, we reject the hypothesis that the US wage distribution did not change at all. Second, we test whether growth was distribution-neutral over the period, i.e. whether $GIC(\tau) = \gamma_{AGR}$. In this test we have an estimate ($\hat{\tau}$) under the null hypothesis and apply the \bar{V}_{1n} and \bar{V}_{2n} tests. Again, we strongly

Figure 1: *GIC* US 1995–2007



reject the null hypothesis, which is in line with the heterogeneity observed across quantiles in Figure 1.

The second interesting finding from our analysis is that the counterfactual growth incidence curve, GIC^* , lies everywhere between the no-growth line at zero, and the actual GIC , and its shape is very similar to that of the latter. This implies that both changes in (broadly defined) economic structure - encompassing changes in returns to observed worker attributes, as well as changes in both the distribution of and returns to unobserved characteristics - and changes in the joint distribution of age, gender and education contributed to the modest increase in US wages during the study period. Since the GIC^* is also flat until $\tau = 0.8$ or thereabouts, and then sharply increasing, we can conclude that the rise in wage inequality is not driven by changes in the gender, age and educational make-up of the workforce. It is driven instead by changes in economic structure and by their impact on the remuneration structure of various worker attributes.

This finding is confirmed by an inspection of the wage inequality measures for the US counterfactual distribution, $F_{Y(1)|T}(\cdot|0)$, in Table 3 above. All five measures lie strictly between the actual wage inequality levels in 1995 and 2007, but are all much closer to the higher 2007 levels. Taking the mean log deviation as an example, the decomposition indicates that changes in economic structure between 1995 and 2007 shifted the measure from 0.218 in 1995 to 0.250. Changes in the joint distribution of covariates - i.e. the age, gender and

Table 4: Kolmogorov-Smirnov (KS) and Cramér-von Mises (CVM) Tests – US

Null Hypothesis	KS	Critical Values			CVM	Critical Values		
		1%	5%	10%		1%	5%	10%
No Effect: $GIC(\tau) = 0$	0.378	0.067	0.0601	0.057	15.677	1.661	1.379	1.184
Mean Effect: $GIC(\tau) = \gamma_{AGR}$	0.306	0.058	0.055	0.051	7.689	1.333	1.252	1.093
No Effect: $GIC^*(\tau) = 0$	0.205	0.080	0.069	0.059	6.590	2.235	1.849	1.694
Mean Effect: $GIC^*(\tau) = \gamma_{AGR^*}$	0.204	0.075	0.063	0.054	5.648	1.472	1.269	1.186

educational make up of the full-time labor force - account only for the residual change from 0.250 to 0.254. Formal tests, also presented in Table 4, confirm that the $GIC^*(\tau)$ was neither constant nor distribution neutral over the period. We first test whether $GIC^*(\tau) = 0$. The results indicate rejection of the null at 1% level for both the KS and CVM tests. And when we test distribution neutrality of growth conditional on the joint distribution of covariates, $GIC^*(\tau) = \gamma_{AGR^*}$ the null is again rejected at all reported levels of significance.

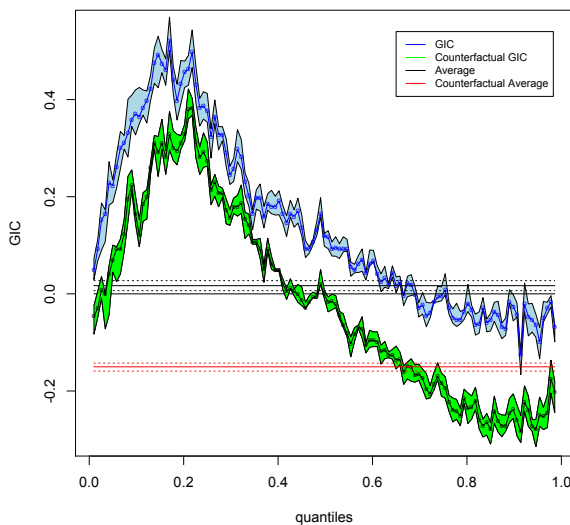
4.2.2 Brazil

The results for the Brazilian GIC and $GIC^*(\tau)$ for 1995-2007 are displayed in Figure 2. As before, the blue line displays the actual GIC , and the dashed black line denotes the growth rate of mean wages, γ_{AGR} . The green line displays the counterfactual growth incidence curve, GIC^* , and the dashed red line shows its corresponding mean effect, γ_{AGR^*} .

Remarkably, there was even less growth in average wages for full-time workers in Brazil than in the US over this period. Cumulative growth in real wages was a paltry 1.7% - a tenth of the US rate.¹¹ However, the distribution of that growth was completely different from the US case. Brazil's GIC rises sharply up until the first quintile, at which point in the distribution wages grew by 40% or more over the period. The GIC is then downward sloping from $\tau = 0.2$ to $\tau = 1.0$. It crosses the x-axis near the 7th decile, and is negative thereafter. This growth pattern is consistent with a substantial decline in wage inequality among full-time workers, as shown in Table 3. Whereas all five inequality indices reported rose for the US, all five declined for Brazil. The Gini coefficient fell by almost four points,

¹¹It is quite likely that this dismal performance is due, at least in part, to a composition effect. Ferreira, Firpo, and Messina (2016) report that formal employment in Brazil rose by a fifth, from 48 to 58% of the labor force, between 1995 and 2012. While not strictly the same, formal employment is highly correlated with full-time status. The same authors also report that the formalization of labor contracts was more common among lower earners. Such a process is likely to lower average earnings in that sample through a composition effect.

Figure 2: *GIC* Brazil 1995–2007



and the mean log deviation, which is more sensitive to income gaps at the bottom of the distribution, lost almost 20% of its initial value.

This pro-poor pattern is also evident in the fact that the average growth rate across quantiles was 13.8% - higher than in the US - despite a near stagnant average wage. Unsurprisingly, then, both the constant distribution and the distribution-neutral growth hypotheses are resoundingly rejected at 1% level of significance for Brazil as well, in both the Kolmogorov-Smirnov and Cramér-von Mises tests. This can be seen in Table 5, which collects the results for the KS and CVM tests (V_{1n} and V_{2n} , respectively).

As in the US case, the counterfactual growth incidence curve lies everywhere below the *GIC*, and has a very similar shape. This parallelism suggests that the main drivers of distributional heterogeneity - which in this case were highly equalizing - belong to the realm of changes in economic structure, affecting remuneration patterns and unobserved worker characteristics. One plausible such candidate driver was the sustained rise in Brazil's minimum wage over this period, which is both consistent with the shape of the *GIC*, and with earlier findings in the literature (e.g. Ferreira, Firpo, and Messina (2016)). Changes in the joint distribution of observed attributes - gender, age and education - on the other hand, had roughly equi-proportional effects across the distribution. These effects were generally positive - i.e. wage-increasing - as one would expect from rising experience and educational levels.

Table 5: Kolmogorov-Smirnov (KS) and Cramér-von Mises (CVM) Tests – Brazil

Null Hypothesis	KS	Critical Values			CVM	Critical Values		
		1%	5%	10%		1%	5%	10%
No Effect: $GIC(\tau) = 0$	0.522	0.0648	0.058	0.055	20.187	1.267	1.151	1.027
Mean Effect: $GIC(\tau) = \gamma_{AGR}$	0.504	0.066	0.062	0.057	19.395	2.024	1.686	1.532
No Effect: $GIC^*(\tau) = 0$	0.383	0.081	0.066	0.059	20.268	1.112	1.045	1.018
Mean Effect: $GIC^*(\tau) = \gamma_{AGR^*}$	0.533	0.090	0.067	0.055	22.666	1.423	1.264	1.178

Once again, this finding is consistent with the inequality measures for the Brazilian counterfactual distribution, reported in the last column of Table 3. These are all lower than the actual inequality values in both 1995 and 2007, suggesting that the observed decline in inequality was due entirely to changes in economic structure. This may well reflect both the effects of a rising minimum wage and the decline in the economy-wide skill premium, as discussed earlier in the literature (see, e.g. Barros et al, 2010). The effect of changes in the observed composition of the labor force was actually to partly offset those declines, through a mildly unequalizing effect of the second term of the decomposition.¹² In terms of formal inference, as should be expected from Figure 2 and the above discussion, both null hypotheses (constant distribution and the distribution-neutral growth) are rejected at the 1% level of significance for both KS and CVM tests. See Table 5.

A comparison of results suggests that the 1995-2007 period saw very different distributional dynamics for real hourly wages among full-time employees across the two countries. Growth in average wages was muted in both countries; and almost zero in Brazil. But such an aggregated description misses important differences in the distribution of that growth: whereas wages were growing at less than 1% per year in the US for all but the top fifth of workers (who experienced much faster increases), Brazil saw relatively rapid wage growth for the bottom half of the distribution, while wages were actually falling for the top fourth. As a result, wage inequality rose in the US and fell markedly in Brazil.

Despite these very disparate headline stories, there were similarities too. In both cases, changes in the observed composition of the labor force - notably higher levels of education and experience - contributed to wage growth, and did so roughly equi-proportionately across

¹²The unequalizing effect of educational expansions when returns are (artificially) held constant is not a novel finding. Bourguignon, Ferreira, and Lustig (2005) refer to this as the 'paradox of progress' and explain that it reflects the generally observed convexity of returns to schooling. As workers become more educated, mass in the schooling distribution shifts to ranges where returns are steeper, and inequality rises.

the distribution. In other words, changes in the joint distribution of \mathbf{X} were not responsible for the sharp movements in inequality in either country. Those changes were attributable almost entirely to changes in the distribution of wages conditional on those observables, interpreted here broadly as changes in economic structure.

5 Conclusion

The recent rise in interest in inequality within the economics profession has not been accompanied by a corresponding ability to properly identify the sources of changes in income or wage distributions. The development of the growth incidence curve (*GIC*) by Ravallion and Chen (2003) has spurred a wave of descriptive studies of the distributional characteristics of economic growth, across many countries and time periods. Hitherto, however, the precise requirements for identification and inference using the *GIC* had not been formally established.

This paper fills that gap by writing the growth incidence curve as a functional of the vector formed by the quantiles of potential outcomes, where treatment assignment is formally replaced by time assignment. We establish the conditions under which both actual and counterfactual growth incidence curves are identified, and propose a simple semi-parametric procedure that allows for the estimation of the *GIC* with no need for restrictive functional form assumptions on the relationship between income and its covariates. We establish the asymptotic properties of these estimators, and propose practical inference procedures for general functions of the quantile potential outcome. Statistical inference procedures are developed uniformly over the set of quantiles \mathcal{T} . We propose testing for general hypotheses and consider both the Kolmogorov-Smirnov and the Cramér-von Mises type statistics. Since the parameter of interest is infinite dimensional, for practical inference, we compute critical values using a bootstrap method. We provide sufficient conditions under which the bootstrap is valid, and discuss an algorithm for its practical implementation.

Finally, we use the proposed methods to estimate the actual and counterfactual growth incidence curves for the US and Brazil, during the 1995-2007 period. The results document important heterogeneity across the quantiles of the income distribution in both growth processes. Neither country had a constant income distribution over that period, and neither growth process was distribution-neutral. Growth in average wages was disappointing in both countries, particularly in Brazil. But these averages hide very different distributional

pictures: Wage stagnation was observed in the US for the bottom 80% of the distribution, while the top fifth, and particularly the top tenth, grew much more rapidly. Conversely, wages rose rapidly below the median in Brazil, and actually fell for the top 25% or so of the distribution. As a result, inequality fell substantially in Brazil and rose in the United States. In both cases, changes in economic structure, rather than in the observed make-up of the labor force, were responsible for changing inequality.

References

- ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings,” *Econometrica*, 70, 91–117.
- ANGRIST, J., V. CHERNOZHUKOV, AND I. FERNANDEZ-VAL (2006): “Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure,” *Econometrica*, 74, 539–563.
- BARROS, R., M. DE CARVALHO, S. FRANCO, AND R. MENDONCA (2010): “Markets, the State, and the Dynamics of Inequality in Brazil,” in *Declining Inequality in Latin America: A Decade of Progress?*, ed. by L. F. Lopez-Calva, and N. Lustig. Washington, DC: Brookings Institution Press.
- BELLONI, A., V. CHERNOZHUKOV, AND I. FERNANDEZ-VAL (2011): “Conditional Quantile Processes Based on Series or Many Regressors,” Working Paper, Boston University.
- BESLEY, T., AND L. J. CORD (2007): *Delivering on the Promise of Pro-Poor Growth*. World Bank and Palgrave MacMillan, Washington DC.
- BITLER, M. P., J. B. GELBACH, AND H. W. HOYNES (2006): “What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments,” *American Economic Review*, 96, 988–1012.
- BOURGUIGNON, F., F. H. G. FERREIRA, AND P. G. LEITE (2008): “Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions,” *Journal of Economic Inequality*, 6, 117–148.
- BOURGUIGNON, F., F. H. G. FERREIRA, AND N. LUSTIG (2005): *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*. World Bank and Oxford University Press, Washington, DC.
- CATTANEO, M. (2010): “Efficient Semiparametric Estimation of Multi-Valued Treatment Effects under Ignorability,” *Journal of Econometrics*, 155, 138–154.
- CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): “Estimation of Semiparametric Models When the Criterion Function is not Smooth,” *Econometrica*, 71, 1591–1608.

- CHERNOZHUKOV, V., AND I. FERNANDEZ-VAL (2005): “Subsampling Inference on Quantile Regression Processes,” *Sankhya*, 67, 253–276.
- CHERNOZHUKOV, V., AND C. HANSEN (2005): “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73, 245–261.
- DEHEJIA, R., AND S. WAHBA (1999): “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs,” *Journal of the American Statistical Association*, 94, 1053–1062.
- DINARDO, J., N. M. FORTIN, AND T. LEMIEUX (1996): “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach,” *Econometrica*, 64, 1001–1044.
- DOKSUM, K. (1974): “Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case,” *The Annals of Statistics*, 2, 267–277.
- DONALD, S. G., D. A. GREEN, AND H. J. PAARSCH (2000): “Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates,” *Review of Economic Studies*, 67, 609–633.
- DONALD, S. G., AND Y.-C. HSU (2014): “Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models,” *Journal of Econometrics*, 178, 383–397.
- ESSAMA-NSSAH, B., S. PAUL, AND L. BASSOLE (2013): “Accounting for Heterogeneity in Growth Incidence in Cameroon Using Recentered Influence Function Regression,” ECINEQ Working Paper no. 289.
- FAN, Y., AND S. S. PARK (2010): “Sharp Bounds on the Distribution of the Treatment Effects and Their Statistical Inference,” *Econometric Theory*, 26, 931–951.
- FERREIRA, F. H. G. (2012): “Distributions in Motion: Economic Growth, Inequality, and Poverty Dynamics,” in *Oxford Handbook of the Economics of Poverty*, ed. by P. Jefferson. Oxford: Oxford University Press.

- FERREIRA, F. H. G., S. FIRPO, AND J. MESSINA (2016): “Understanding Recent Earnings Inequality Dynamics in Brazil,” in *New Order and Progress: Development and Democracy in Brazil*, ed. by B. R. Schneider. New York: Oxford University Press.
- FIRPO, S. (2007): “Efficient Semiparametric Estimation of Quantile Treatment Effects,” *Econometrica*, 75, 259–276.
- FIRPO, S., AND C. PINTO (2015): “Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures,” *Journal of Applied Econometrics*, 31, 457–486.
- FLORES, C. A. (2007): “Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment,” mimeo.
- GALVAO, A. F., AND L. WANG (2015): “Uniformly Semiparametric Efficient Estimation of Treatment Effects with a Continuous Treatment,” *Journal of the American Statistical Association*, 110, 1528–1542.
- GUTENBRUNNER, C., AND J. JURECKOVA (1992): “Regression Rank Scores and Regression Quantiles,” *The Annals of Statistics*, 20, 305–330.
- HECKMAN, J., H. ICHIMURA, J. SMITH, AND P. TODD (1998): “Characterizing Selection Bias Using Experimental Data,” *Econometrica*, 66, 1017–1098.
- HIRANO, K., G. IMBENS, AND G. RIDDER (2003): “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,” *Econometrica*, 71, 1161–1189.
- HIRANO, K., AND G. W. IMBENS (2004): “The Propensity Score with Continuous Treatment,” in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, ed. by A. Gelman, and X.-L. Meng. Wiley.
- JUHN, C., K. M. MURPHY, AND B. PIERCE (1993): “Wage Inequality and the Rise in Returns to Skill,” *Journal of Political Economy*, 101, 410–442.
- KOENKER, R., AND G. W. BASSETT (1978): “Regression Quantiles,” *Econometrica*, 46, 33–49.
- KOENKER, R., S. LEORATO, AND F. PERACCHI (2013): “Distributional vs. Quantile Regression,” CEIS Tor Vergata, Research Paper Series.

- KOENKER, R., AND J. A. F. MACHADO (1999): “Godness of Fit and Related Inference Processes for Quantile Regression,” *Journal of the American Statistical Association*, 94, 1296–1310.
- KOENKER, R., AND Z. XIAO (2002): “Inference on the Quantile Regression Process,” *Econometrica*, 70, 1583–1612.
- KOPCZUK, W., E. SAEZ, AND J. SONG (2010): “Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937,” *Quarterly Journal of Economics*, 125, 91–128.
- KOSOROK, M. (2008): *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, NY.
- LEHMANN, E. L. (1974): *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, CA.
- LINTON, O., E. MAASOUMI, AND Y.-J. WHANG (2005): “Consistent Testing for Stochastic Dominance under General Sampling Schemes,” *Review of Economic Studies*, 72, 735–765.
- MISHEL, L., J. BIVENS, E. GOULD, AND H. SHIERHOLZ (2012): *The State of Working America*. 12th Edition. An Economic Policy Institute book. Cornell University Press, Ithaca, N.Y.
- NEWBY, W. K. (1997): “Convergence Rates and Asymptotic Normality for Series Estimators,” *Journal of Econometrics*, 79, 147–168.
- QU, Z., AND J. YOON (2015): “Nonparametric Estimation and Inference on Conditional Quantile Processes,” *Journal of Econometrics*, 185, 1–19.
- RAVALLION, M., AND S. CHEN (2003): “Measuring Pro-poor Growth,” *Economics Letters*, 78, 93–99.
- RUBIN, D. (1977): “Assignment to Treatment Group on the Basis of a Covariate,” *Journal of Educational Statistics*, 2, 1–28.
- VAN DER VAART, A. (2002): “Infinite-dimensional Z-Estimators,” in *Lectures on Probability Theory and Statistics*, ed. by P. Bernard. Berlin, Springer Verlag.

VAN DER VAART, A., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag Press, New York, New York.

——— (2007): “Empirical Processes Indexed by Estimated Functions,” *IMS Lecture Notes Monograph Series, Asymptotics: Particles, Processes and Inverse Problems*, 55, 234–252.

6 Appendix

6.1 Proofs of the main results

This appendix collects the proofs of the results given in the text. To demonstrate Theorems 1, 2, and 3 below we make use of Lemmas 4, 5, and 6, respectively, given in the Online Supplemental Appendix. These lemmas establish, respectively, uniform consistency, weak convergence, and validity of the bootstrap for generic Z-estimators with possibly non-smooth functions and a nuisance parameter, when both the parameter of interest and the nuisance parameter are possibly infinitely dimensional. The results allow for the case of profiled nonparametric estimator, i.e., it depends on the parameters.

For clarity, the demonstrations make use of the superscript zero to denote the true parameters.

Proof of Theorem 1. The general result for consistency of Z-estimator is given in **Lemma 4** in the Online Supplemental Appendix. To prove the result we apply the lemma to our continuous treatment model with $\theta_0 = q^0(\cdot)$, $h_0 = w^0(\cdot)$, $\mathbb{Z}(\theta, h)(\tau) = \mathbb{E}\psi_{q,w,\tau}$, and $Z(\theta, h)(\tau) = \mathbb{E}\psi_{q,w,\tau}$, where $\psi_{q,w,\tau} = m(y, q(\tau); \tau)w(\mathbf{x}) = (\tau - \mathbf{1}\{Y < q(\tau)\})w(\mathbf{x})$. Notice that $w(T, \mathbf{x})$, and since $T = \{0, 1\}$ we write $w(\mathbf{x}) = w(T, \mathbf{x})$ in the demonstrations.

In this case, $\Theta = \mathcal{L} = \ell^\infty(\mathcal{T})$ and $\|\cdot\|_\Theta = \|\cdot\|_{\mathcal{L}} = |\cdot|_\infty$, while $\mathcal{H} = \Pi$, a function class with domain $\{0, 1\} \times \mathcal{X}$, and $\|\cdot\|_{\mathcal{H}} = \|\cdot\|_\Pi = \sup_{\mathbf{x} \in \mathcal{X}} |\cdot| = |\cdot|_\infty$. For any $\delta > 0$, $\Pi_\delta = \{w \in \Pi : |w - w^0|_\infty < \delta\}$.

To establish the result we verify the conditions of Lemma 4. Thus, under **QC.I–QC.III** we can check the general conditions **C.1–C.5** in the Supplemental Appendix. Condition **C.1** is satisfied by the computational properties of quantile regression estimator of Theorem 3.3 of Koenker and Bassett (1978) and conditions **QC.II** and **QC.III** such that we have

$$\begin{aligned} |\mathbb{E}[(\tau - \mathbf{1}\{Y < \hat{q}_t(\cdot)\})\hat{w}(\mathbf{X})]| &\leq \text{const} \cdot \sup_{i \leq n} \frac{\hat{w}_t(\mathbf{X}_i)}{n} \\ &\leq \text{const} \cdot \frac{\|\hat{w}_t(\mathbf{X})\|_\Pi}{n} = \text{const} \cdot \frac{\|w_t^0(\mathbf{x})\|_\Pi + o_p(1)}{n} = O_{p^*}(1/n). \end{aligned}$$

Condition **C.2** holds by condition **QC.I**.

We now show that condition **C.3**, the continuity of $\mathbb{E}[m(Y; q_t(\tau))w_t(\mathbf{X})]$ at w_t^0 uniformly over $q_t(\tau) \in \ell^\infty(\mathcal{T})$, is satisfied. For any $\|w_t - w_t^0\|_\infty \leq \delta$, which is equivalent to

$\sup_{\tau \in \mathcal{T}} \sup_{\mathbf{x} \in \mathcal{X}} |w_t(\mathbf{x}) - w_t^0(\mathbf{x})| \leq \delta$, we have

$$\begin{aligned} & \sup_{\tau \in \mathcal{T}} \sup_{\mathbf{x} \in \mathcal{X}} |\mathbb{E}[m(Y, q_t(\tau); \tau)w_t(\mathbf{X})] - \mathbb{E}[m(Y, q_t(\tau); \tau)w_t^0(\mathbf{X})]| \\ &= \sup_{\tau \in \mathcal{T}} \sup_{\mathbf{x} \in \mathcal{X}} |\mathbb{E}[m(Y, q_t(\tau); \tau)(w_t(\mathbf{X}) - w_t^0(\mathbf{X}))]| \leq \sup_{\tau \in \mathcal{T}} |\mathbb{E}[m(Y, q_t(\tau); \tau)]| \delta. \end{aligned}$$

Therefore, condition **C.3** is satisfied because $\tau - \mathbf{1}\{y < q_t(\tau)\}$ is a bounded function.

Note that the functional class $\{\psi_{q,w,\tau} = (\tau - \mathbf{1}\{Y < q(\tau)\})w(\mathbf{X}), q \in \Theta, w \in \Pi, \tau \in \mathcal{T}\}$ is formed as $(\mathcal{T} - \mathcal{F})w(\mathbf{X})$, where $\mathcal{F} = \mathbf{1}\{Y < q(\tau)\}$ is a VC subgraph class and hence a bounded Donsker class. Hence $\mathcal{T} - \mathcal{F}$ is also bounded Donsker, and using assumption **QC.II**, $(\mathcal{T} - \mathcal{F})w(\mathbf{X})$ is therefore Donsker with a square integrable envelope $2 \max_t |w(\mathbf{X})|_t$, by Theorem 2.10.6 in Van der Vaart and Wellner (1996). The stochastic equicontinuity then is a part of being Donsker, which implies condition **C5S** which in turn implies **C.5**.

Hence, all the conditions of Lemma 4 are satisfied. \square

Proof of Theorem 2. To establish the result we apply Lemma 5 in the Online Supplemental Appendix and we verify its conditions.

Condition **G.1** was verified in the proof of Theorem 1. For condition **G.2**, note that

$$\begin{aligned} & |\mathbb{E}[(\tau - \mathbf{1}\{Y \leq q_t(\cdot)\})w_t^0(\mathbf{X})] - \mathbb{E}[(\tau - \mathbf{1}\{Y \leq q_t^0(\cdot)\})w_t^0(\mathbf{X})] + \mathbb{E}[w_t^0(\mathbf{X})f_Y(q_t^0)(q_t(\cdot) - q_t^0(\cdot))]|_\infty \\ &= |\mathbb{E}[\{\mathbf{1}\{Y \leq q_t^0(\cdot)\} - \mathbf{1}\{Y \leq q_t(\cdot)\} + f_Y(q_t^0)(q(\cdot) - q_t^0(\cdot))\}w_t^0(\mathbf{X})]|_\infty \\ &\asymp |\mathbb{E}[\{\mathbf{1}\{Y \leq q_t^0(\cdot)\} - \mathbf{1}\{Y \leq q_t(\cdot)\} + f_Y(q_t^0)(q_t(\cdot) - q_t^0(\cdot))\}]|_\infty M_w \\ &= |F_Y(q_t^0(\cdot)) - F_Y(q_t(\cdot)) + f_Y(q_t^0)(q_t(\cdot) - q_t^0(\cdot))|_\infty M_w = o(|q_t(\cdot) - q_t^0(\cdot)|_\infty). \end{aligned}$$

Now we verify condition **G.3**. To find the pathwise derivative of $Z(q_t, w_t^0)$ with respect to w_t , we conduct the following calculations. For any \bar{w}_t such that $\{w_t^0 + \alpha(\bar{w}_t - w_t^0) : \alpha \in [0, 1]\} \subset \Pi$,

$$\frac{\mathbb{E}[m(Y, q_t; \tau)(w_t^0 + \alpha(\bar{w}_t - w_t^0))] - \mathbb{E}[m(Y, q_t; \tau)w_t^0]}{\alpha} = \mathbb{E}[m(Y, q_t; \tau)(\bar{w}_t - w_t^0)]$$

and has the limit $\mathbb{E}[m(Y, q_t; \tau)(\bar{w}_t - w_t^0)]$ as $\alpha \rightarrow 0$. Therefore $Z_2(q_t, w_t^0)[w_t - w_t^0] = \mathbb{E}[m(Y, q_t; \tau)(w_t - w_t^0)]$ in all directions $[w_t - w_t^0] \in \Pi$. Condition **G.3.1** is satisfied by

noting that

$$|\mathbb{E}[m(Y, q_t(\cdot), \cdot)w_t^0(\mathbf{X})] - \mathbb{E}[m(Y, q_t(\cdot), \cdot)w_t(\mathbf{X})] - \mathbb{E}[m(Y, q_t(\cdot); \cdot)(w_t - w_t^0)(\mathbf{X})]|_\infty = 0.$$

And condition **G.3.2** is verified by

$$\begin{aligned} & |\mathbb{E}[m(Y, q_t(\cdot); \cdot)(w_t - w_t^0)(\mathbf{X})] - \mathbb{E}[m(Y, q_t^0(\cdot); \cdot)(w_t - w_t^0)(\mathbf{X})]|_\infty \\ &= |\mathbb{E}[m(Y, q_t(\cdot); \cdot) - m(Y, q_t^0(\cdot); \cdot)(w_t - w_t^0)(\mathbf{X})]|_\infty \\ &\leq |\mathbb{E}[m(Y, q_t(\cdot); \cdot)] - \mathbb{E}[m(Y, q_t^0(\cdot); \cdot)]|_\infty o(1) = \delta_n o(1), \end{aligned}$$

where the last equality follows because the distribution function of Y is continuous.

Condition **G.4** is automatically satisfied by **QG.III**. Now we check condition **G.5**. Note that $\{\psi_{q,w,\tau} : q \in \ell^\infty(\mathcal{T}), w \in \Pi, \tau \in \mathcal{T}\}$ is Donsker. This follows because by **QG.I** the bracketing number of Π by Corollary 2.7.4 in van der Vaart and Wellner (1996) is finite, thus Π is Donsker with a constant envelope. The class \mathcal{F} is Donsker by exploiting the monotonicity and boundedness of indicator function and bounded density condition assumed in **QC.I**. Finally, the result follows because the class is formed by taking products and sums of bounded Donsker classes \mathcal{F} , Π , and \mathcal{T} , which is Lipschitz over $(\mathcal{F} \times \Pi \times \mathcal{T})$. Hence by Theorem 2.10.6 in van der Vaart and Wellner (1996) $\{\psi_{q,w,\tau}\}$ is Donsker and we have that **G.5'** is satisfied by Lemma 3.3.5 of van der Vaart and Wellner (1996). Therefore, we obtain condition **G.5** by condition **G.1** and inequality (4) in Lemma 5 in the Online Supplemental Appendix.

Finally, condition **G.6** holds by **QG.II**. Hence, all the conditions of Theorem 2 are satisfied. \square

Proof of Corollary 1. The proof follows directly from the result in Theorem 2, which establish Donsker properties, and therefore tightness, for each element of the vector. By noticing that marginal tightness implies joint tightness, and from joint finite-dimensional asymptotic normality, the result follows. \square

Proof of Lemma 1. The proof follows from the result in Theorem 2 and Corollary 1 and the functional delta method. Corollary 1 implies the weak convergence result, $\sqrt{n}(\widehat{Q}(\tau) -$

$Q^0(\tau) \rightsquigarrow \mathbb{G}(\tau)$. From the assumptions and the differentiability condition in **QG.IV** of $h(q_t(\tau))$ at q_t Theorem 3.9.5 in van der Vaart and Wellner (1996) applies and the result follows. \square

Proof of Corollary 2. The proof follows directly from the result in Theorem 2, Lemma 1, and the Hadamard differentiability of GIC and GIC^* . \square

Proof of Lemma 2. The assertion holds by Corollary 1, Lemma 1, and the continuous mapping theorem. \square

Proof of Corollary 3. The assertion holds by Lemma 2 and the continuous mapping theorem. \square

Proof of Lemma 3. The assertion holds by Corollary 1, Lemma 1, and the continuous mapping theorem. \square

Proof of Corollary 4. The assertion holds by Lemma 3 and the continuous mapping theorem. \square

Proof of Theorem 3. This theorem is a restatement of the Lemma 3 in the Supplemental Appendix. \square

6.2 Semiparametric efficiency of the two-step estimator

In this section, we establish the uniform semiparametric efficiency of the two-step estimator. We first calculate the efficient influence function of the parameter $q_t(\tau)$ in the following semiparametric model

$$\mathcal{F} = \{F_{q,w} : q \in \ell^\infty(\mathcal{T}), w \in \Pi\},$$

where F_{q^0, w^0} is the distribution function of the observed data. Then, we provide sufficient conditions under which the proposed two-step estimator is uniformly semiparametric efficient.

Proposition 1. *Suppose $\Gamma_0(\tau) := \frac{\partial \mathbb{E}[m(Y(t), q^0(\tau); \tau)]}{\partial \beta(\tau)}$ exists for $t = \{0, 1\}$. For each $\tau \in \mathcal{T}$, the efficient influence function of the parameter $q(\tau)$ is*

$$\Psi_q(y, t, \mathbf{x}, \tau) = -\Gamma_0^{-1}(t, \tau) \psi(y, \mathbf{x}, t, q^0(\tau), w^0, e^0),$$

where $\psi(y, \mathbf{x}, t, q^0(\tau), w^0, e^0) = m(y, q^0(\tau); \tau) w^0(\mathbf{x}) - e^0(\mathbf{x}, q^0(\tau)) (w^0(\mathbf{x}) - 1)$ with $e^0(\mathbf{x}, q(\tau)) = \mathbb{E}[m(Y, q^0(\tau); \tau) | \mathbf{X} = \mathbf{x}]$.

Proof. The proof is given in Theorem 3 of Firpo (2007). □

Based on the efficient influence function of $q(\tau)$, we show that the two-step estimator is uniformly semiparametric efficient provided the following condition

$$\mathbf{E}. \quad \sqrt{n} \mathbb{E}[m(Y, q^0(\tau); \tau) \widehat{w}(\mathbf{X})] = \sqrt{n} \mathbb{E}[\psi(Y, \mathbf{X}, t, q^0(\tau), w^0, e^0)] + o_p(1).$$

Condition **E** is critical to the efficiency of the two-step estimator, and it is similar to its corresponding condition for the multi-valued model is condition (4.2) of Cattaneo (2010).

Theorem 4. *Assume that the conditions of Theorem 2 in the main text and condition **E** hold. Then the two-step estimator is uniformly semiparametric efficient.*

This result guarantees that the two-step estimator is uniformly semiparametric efficient. Hypothesis testings based on this estimator are expected to be optimal.

Proof of Theorem 4. We first verify that $\sqrt{n} \mathbb{E}[\psi(Y_i, \mathbf{X}_i, t, q_t^0, w_t^0, e^0)]$ converges weakly in $\ell^\infty(\mathcal{T})$. Proceeding in the exact same way as in the proof of Theorem 2, conditions **QC.I** and **QG.I** imply **G.5** (in the Online Supplemental Appendix), and hence $\psi_t = \psi(y, \mathbf{x}, t, q_t^0, w_t^0, e^0)$ is Donsker, which in turn implies the weak convergence.

The uniform semiparametric efficiency follows from the weak convergence above and the pointwise semiparametric efficiency (Theorem 3 in Firpo (2007)) by Theorem 18.9 of Kosorok (2008).

Now we verify that the formula in condition **QG.II** equals the left hand-side of condition **E.**, which implies that the influence function of the two-step estimator is efficient. Recall that $m(Y(t), q_t(\tau); \tau) = \tau - \mathbf{1}\{Y(t) < q_t(\tau)\}$. To this end, we begin with the formula in condition **QG.II**.

$$\begin{aligned}
& \sqrt{n}\mathbb{E}[(\tau - \mathbf{1}\{Y < q_t^0(\tau)\})(w_t(\mathbf{X}) - w_t^0(\mathbf{X}))]|_{w=\widehat{w}} \\
&= \sqrt{n}(\mathbb{E}m(Y, q_t^0(\tau); \tau)(w_t(\mathbf{X}) - w_t^0(\mathbf{X}))|_{w=\widehat{w}} \\
&= \sqrt{n}(\mathbb{E}m(Y, q_t^0(\tau); \tau)(\widehat{w}_t(\mathbf{X}) - w_t^0(\mathbf{X})) \\
&= \sqrt{n}(\mathbb{E}m(Y, q_t^0(\tau); \tau)\widehat{w}_t(\mathbf{X})),
\end{aligned}$$

where the first equality uses the definition of $m(\cdot)$, and the second equality follows by condition **G5'** which in turn was verified in the proof of Theorem 2. \square

6.3 Estimation of weights $w(\mathbf{X})$

The estimation of the nuisance parameter in the first step is very important for practical implementation of the proposed methods. We have been assuming that the estimator \widehat{w}_t of the nuisance parameter w_t^0 satisfies various conditions (**QC.III** and **QG.I–QG.III**). In this section we discuss the estimation of the weights for QTE, QTT, and QTU.

Recall that $w_1(\mathbf{X}) = \frac{T}{p(\mathbf{X})}$, $w_0(\mathbf{X}) = \frac{1-T}{1-p(\mathbf{X})}$, $w_{A1}(\mathbf{X}) = \frac{T}{p}$, $w_{A0}(\mathbf{X}) = \frac{1-T}{p} \frac{p(\mathbf{X})}{1-p(\mathbf{X})}$, $w_{B1}(\mathbf{X}) = \frac{T}{1-p} \frac{1-p(\mathbf{X})}{p(\mathbf{X})}$ and $w_{A0}(\mathbf{X}) = \frac{1-T}{1-p}$. The estimators are defined by the plug-in method as following: $\widehat{w}_1(\mathbf{X}) = \frac{T}{\widehat{p}(\mathbf{X})}$, $\widehat{w}_0(\mathbf{X}) = \frac{1-T}{1-\widehat{p}(\mathbf{X})}$, $\widehat{w}_{A1}(\mathbf{X}) = \frac{T}{\widehat{p}}$, $\widehat{w}_{A0}(\mathbf{X}) = \frac{1-T}{\widehat{p}} \frac{\widehat{p}(\mathbf{X})}{1-\widehat{p}(\mathbf{X})}$, $\widehat{w}_{B1}(\mathbf{X}) = \frac{T}{1-\widehat{p}} \frac{1-\widehat{p}(\mathbf{X})}{\widehat{p}(\mathbf{X})}$ and $\widehat{w}_{A0}(\mathbf{X}) = \frac{1-T}{1-\widehat{p}}$. Therefore, the important pieces for estimation are the conditional probability of being treated, $p(\mathbf{X}) = Pr[T = 1|\mathbf{X} = \mathbf{x}]$, and the unconditional $p = Pr[T = 1]$. The latter can be estimated by its sample counterpart, that is, $\widehat{p} = \frac{\sum_{i=1}^n T_i}{n}$. For the former, we follow Firpo (2007). Following the propensity score estimation strategy employed by HIR, we use a logistic power series approximation, i.e., a series of functions of \mathbf{X} is used to approximate the log-odds ratio of the propensity score. The log-odds ratio of $p(\mathbf{x})$ is equal to $\log(p(\mathbf{x})/(1-p(\mathbf{x})))$. These functions are chosen to be polynomials of \mathbf{x} and the coefficients that correspond to those functions are estimated by a pseudo-maximum likelihood method.

Start by defining $H_K(\mathbf{x}) = [H_{K,j}(\mathbf{x})]$ ($j = 1, \dots, K$), a vector of length K of polynomial functions of $\mathbf{x} \in \mathcal{X}$ satisfying the following properties: (i) $H_K : \mathcal{X} \rightarrow \mathbb{R}^K$; and (ii) $H_{K,1}(\mathbf{x}) = 1$. If we want $H_K(\mathbf{x})$ to include polynomials of \mathbf{x} up to the order n , then it is

sufficient to choose K such that $K \geq (n + 1)^r$. In what follows, we will assume that K is a function of the sample size N and grows without bounds as n grows without bounds, that is, $K = K(n) \rightarrow \infty$ as $n \rightarrow \infty$.

Next, the propensity score is estimated. Let $\hat{p}(\mathbf{x}) = L(H_K(\mathbf{x})'\hat{\pi})$, where $L : \mathbb{R} \rightarrow \mathbb{R}$, $L(z) = (1 + \exp(-z))^{-1}$; and

$$\hat{\pi} = \arg \max_w \frac{1}{n} \sum_{i=1}^n (T_i \log(L(H_K(\mathbf{X}_i)'\pi)) + (1 - T_i) \log(1 - L(H_K(\mathbf{X}_i)'\pi))).$$

The asymptotic properties of the logistic power series as discussed in Hirano, Imbens, and Ridder (2003) and Newey (1997). The required conditions (**QC.III** and **QG.I–QG.III**) are satisfied when using the logistic power series estimator. **QG.I** follows directly from the asymptotic properties of estimator. **QG.II** is satisfied by exploiting the monotonicity and boundedness of indicator function and bounded density condition assumed in **QC.I**. Finally, **QC.III** and **QG.III** follow from the mean value theorem.

7 Supplemental Appendix (Online)

This supplement contain two parts. First, it presents results for the asymptotic theory for the generic Z-estimator. Second, we provide Monte Carlo simulations to evaluate the finite sample performance of the proposed methods. The simulations provide evidence that the methods perform well in finite samples. The empirical size of the test approximates the nominal one, and the tests have large empirical power.

7.1 Asymptotic Theory

In this appendix, we establish the asymptotic properties of a generic Z-estimator. More specifically, we describe the model, the regularity conditions, and state the asymptotic results.

In Lemmas 4 and 5 below, we provide verifiable sufficient conditions for general consistency and weak convergence of generic moment restriction estimators (Z-estimators) with possibly non-smooth functions and a nuisance parameter, when both the parameter of interest and the nuisance parameter are possibly infinite dimensional. The results allow for the case where the nonparametric estimator is profiled, i.e., is allowed to depend on the parameters. Lemma 6 establishes the validity of the bootstrap. These general results are used to prove the asymptotic properties of the two-step estimator discussed in the main text. In this general setting, the data need not be independent and identically distributed (i.i.d.). These approaches and results are similar to those in van der Vaart (2002) and van der Vaart and Wellner (2007). While these later works provide high level conditions, we describe simpler verifiable conditions for Z-estimators. The results for the general theory presented here extend those of Chen, Linton, and Van Keilegom (2003) in that the parameter of interest is a Banach valued quantity instead of a Euclidean vector. Moreover, the results extend Theorem 3.3.1 of van der Vaart and Wellner (1996) in that a possibly infinite dimensional nuisance parameter needs to be estimated in the first step.

Let Θ and \mathcal{L} denote Banach spaces, and \mathcal{H} a norm space, with norms $\|\cdot\|_{\Theta}$, $\|\cdot\|_{\mathcal{H}}$, and $\|\cdot\|_{\mathcal{L}}$, respectively. Let $\mathbb{Z}_n : \Theta \times \mathcal{H} \mapsto \mathcal{L}$, $Z : \Theta \times \mathcal{H} \mapsto \mathcal{L}$ be random maps and a deterministic map, respectively. We suppress the dependence of \mathbb{Z} on n for simplicity. The Z-estimator $\hat{\theta}$ is defined as the approximate root of

$$\mathbb{Z}(\theta, \hat{h}) = \mathbf{0},$$

where \widehat{h} is a first step estimator of a possibly infinite dimensional nuisance parameter.

7.2 Consistency

We first derive a general consistency result for a Z-estimator in a Banach space. To obtain the consistency of the generic Z-estimator, we impose the following conditions.

C.1 $\|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} = o_{p^*}(1)$.

C.2 $\|Z(\theta_n, h_0)\|_{\mathcal{L}} \rightarrow 0$ implies $\theta_n \rightarrow \theta_0$ for any sequences $\theta_n \in \Theta$.

C.3 Uniformly in $\theta \in \Theta$, $Z(\theta, h)$ is continuous at h_0 .

C.4 $\|\widehat{h} - h_0\|_{\mathcal{H}} = o_{p^*}(1)$.

C.5 For all sequences $\delta_n \downarrow 0$,

$$\sup_{\theta \in \Theta, \|h - h_0\|_{\mathcal{H}} \leq \delta_n} \frac{\|\mathbb{Z}(\theta, h) - Z(\theta, h)\|_{\mathcal{L}}}{1 + \|\mathbb{Z}(\theta, h)\|_{\mathcal{L}} + \|Z(\theta, h)\|_{\mathcal{L}}} = o_{p^*}(1).$$

Condition **C.1** requires that $\widehat{\theta}$ solves the estimating equation $\|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} = 0$ only asymptotically. Condition **C.2** is an identification of the parameter. Condition **C.3** is a smooth assumption of Z in h only at h_0 . Condition **C.4** requires that the nuisance parameter is consistently estimated. Condition **C.5** is a high level assumption and can be stated in more primitive conditions for specific cases. Further, condition **C.5** is implied by the following uniform convergence condition of \mathbb{Z} to Z .

C5S For any sequences $\delta_n \downarrow 0$,

$$\sup_{\theta \in \Theta, \|h - h_0\|_{\mathcal{H}} \leq \delta_n} \|\mathbb{Z}(\theta, h) - Z(\theta, h)\|_{\mathcal{L}} = o_{p^*}(1).$$

This set of conditions are similar to conditions of Theorem 1 of Chen, Linton, and Van Keilegom (2003).

The following lemma summarizes the consistency of the generic Z-estimator.

Lemma 4. *Suppose that $\theta_0 \in \Theta$ satisfies $Z(\theta_0, h_0) = 0$ with $h_0 \in \mathcal{H}$ and that conditions **C.1–C.5** hold. Then $\|\widehat{\theta} - \theta_0\|_{\Theta} = o_{p^*}(1)$.*

Proof. By condition **C.2**, it suffices to show that $\|Z(\widehat{\theta}, h_0)\|_{\mathcal{L}} = o_{p^*}(1)$. Using the triangle inequality,

$$\|Z(\widehat{\theta}, h_0)\|_{\mathcal{L}} \leq \|Z(\widehat{\theta}, h_0) - Z(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} + \|Z(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} + \|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}}.$$

By conditions **C.3** and **C.4**, $\|Z(\widehat{\theta}, h_0) - Z(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} = o_{p^*}(1)$. By condition **C.1**, $\|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} = o_{p^*}(1)$. In addition,

$$\begin{aligned} \|Z(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} &= o_{p^*}(1) + o_{p^*}(\|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}}) + o_{p^*}(\|Z(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}}) \\ &= o_{p^*}(1) + o_{p^*}(1) + o_{p^*}(\|Z(\widehat{\theta}, h_0)\|_{\mathcal{L}}) + o_{p^*}(1), \end{aligned}$$

where the first equality follows by condition **C.5** and the second equality is a result of conditions **C.1** and **C.3**. Therefore, inequality implies $\|Z(\widehat{\theta}, h_0)\|_{\mathcal{L}} \leq o_{p^*}(1)$ and hence the result. \square

7.3 Weak Convergence

Now we provide a general result of weak convergence for the Z-estimator. For the proof of weak convergence of the Z-estimator, consistency is assumed without loss of generality. Therefore, the parameter space is replaced by $\Theta_\delta \times \mathcal{H}_\delta$ where $\Theta_\delta := \{\theta \in \Theta : \|\theta - \theta_0\|_{\Theta} < \delta\}$ as in Chen, Linton, and Van Keilegom (2003) and $\mathcal{H}_\delta := \{h \in \mathcal{H} : \|h - h_0\|_{\mathcal{H}} < \delta\}$.

Because the parameter spaces are a Banach and a normed space, we need notions of derivatives for maps from a Banach or a normed space to a Banach space. Let Θ and \mathbb{L} denote Banach spaces, and \mathbb{H} a normed space. Fréchet differentiability of a map $\phi : \Theta \mapsto \mathbb{L}$ at $\theta \in \Theta$ means that there exists a continuous, linear map $\phi'_\theta : \Theta \mapsto \mathbb{L}$ with

$$\frac{\|\phi(\theta + h_n) - \phi(\theta) - \phi'_\theta(h_n)\|}{\|h_n\|} \rightarrow 0$$

for all sequences $\{h_n\} \subset \Theta$ with $\|h_n\| \rightarrow 0$ and $\theta + h_n \in \Theta$ for all $n \geq 1$; see, e.g., p. 26 of Kosorok (2008). Pathwise derivative of a map $\varphi : \mathbb{H} \mapsto \mathbb{L}$ at $h \in \mathbb{H}$ in the direction $[\bar{h} - h]$ is

$$\varphi'_h[\bar{h} - h] = \lim_{\varrho \rightarrow 0} \frac{\varphi(h + \varrho(\bar{h} - h)) - \varphi(h)}{\varrho}$$

with $\{h + \varrho(\bar{h} - h) : \varrho \in [0, 1]\} \subset \mathbb{H}$, provided that the limit exists. To obtain the weak limit, we impose the following sufficient conditions.

G.1 $\|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} = o_{p^*}(n^{-1/2})$.

G.2 The map $\theta \mapsto Z(\theta, h_0)$ is Fréchet differentiable at θ_0 with a continuously invertible derivative $Z_1(\theta_0, h_0)$.

G.3 For all $\theta \in \Theta_\delta$ the pathwise derivative $Z_2(\theta, h_0)[h - h_0]$ of $Z(\theta, h_0)$ exists in all directions $[h - h_0] \in \mathcal{H}$. Moreover, for all $(\theta, h) \in \Theta_{\delta_n} \times \mathcal{H}_{\delta_n}$ with a positive sequence $\delta_n = o(1)$:

G.3.1 $\|Z(\theta, h_0) - Z(\theta, h) - Z_2(\theta, h_0)[h - h_0]\|_{\mathcal{L}} \leq c\|h - h_0\|_{\mathcal{H}}^2$ for a constant $c \geq 0$.

G.3.2 $\|Z_2(\theta, h_0)[h - h_0] - Z_2(\theta_0, h_0)[h - h_0]\|_{\mathcal{L}} \leq o(1)\delta_n$.

G.4 The estimator $\widehat{h} \in \mathcal{H}$ with probability tending to one; and $\|\widehat{h} - h_0\|_{\mathcal{H}} = o_{p^*}(n^{-1/4})$.

G.5 For any $\delta_n \downarrow 0$,

$$\sup_{\|\theta - \theta_0\| \leq \delta_n, \|h - h_0\|_{\mathcal{H}} \leq \delta_n} \frac{\|\sqrt{n}(\mathbb{Z} - Z)(\theta, h) - \sqrt{n}(\mathbb{Z} - Z)(\theta_0, h_0)\|_{\mathcal{L}}}{1 + \sqrt{n}\|\mathbb{Z}(\theta, h)\|_{\mathcal{L}} + \sqrt{n}\|Z(\theta, h)\|_{\mathcal{L}}} = o_{p^*}(1).$$

G.6 $\sqrt{n}(Z_2(\theta_0, h_0)[\widehat{h} - h_0] + (\mathbb{Z} - Z)(\theta_0, h_0))$ converges weakly to a tight random element \mathbb{G} in \mathcal{L} .

Condition **G.1** requires $\widehat{\theta}$ to solve the estimating equation only asymptotically. Conditions **G.2** and **G.3** are smoothness conditions for Z . Condition **G.4** is the same as condition (2.4) of Chen, Linton, and Van Keilegom (2003). Conditions **G.5** and **G.6** are high level assumptions, and more primitive conditions are provided for more specific cases. Moreover, condition **G.5** is implied by

G.5' For any $\delta_n \downarrow 0$,

$$\sup_{\|\theta - \theta_0\| \leq \delta_n, \|h - h_0\|_{\mathcal{H}} \leq \delta_n} \|\sqrt{n}(\mathbb{Z} - Z)(\theta, h) - \sqrt{n}(\mathbb{Z} - Z)(\theta_0, h_0)\|_{\mathcal{L}} = o_{p^*}(1).$$

Now we provide a general result for Z -estimators.

Lemma 5. *Suppose that $\theta_0 \in \Theta_\delta$ satisfies $Z(\theta_0, h_0) = 0$, that $\widehat{\theta} = \theta_0 + o_{p^*}(1)$, and that conditions **G.1–G.6** hold. Then,*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \rightsquigarrow Z_1^{-1}(\theta_0, h_0)\mathbb{G}.$$

Proof. The proof is divided in two steps. First, we establish \sqrt{n} -consistency. Second, we establish the weak convergence.

Step 1: \sqrt{n} -consistency

We start the proof by showing that $\widehat{\theta}$ is \sqrt{n} -consistent for θ_0 in Θ . By definition, the Fréchet differentiability of $Z(\theta, h_0)$ implies the existence of a continuous linear map $Z_1(\theta_0, f_0)$ such that

$$\frac{\|Z(\theta, f_0) - Z(\theta_0, f_0) - Z_1(\theta_0, f_0)(\theta - \theta_0)\|_{\mathcal{L}}}{\|\theta - \theta_0\|_{\Theta}} = o(1).$$

By the triangle inequality, it follows

$$\|Z_1(\theta_0, h_0)(\theta - \theta_0)\|_{\mathcal{L}} \leq \|Z(\theta, h_0) - Z(\theta_0, h_0)\|_{\mathcal{L}} + o(\|\theta - \theta_0\|_{\Theta}).$$

Since the derivative $Z_1(\theta_0, h_0)$ is continuously invertible by condition **G.2**, there exists a positive constant c such that $\|Z_1(\theta_0, h_0)(\theta_1 - \theta_2)\|_{\mathcal{L}} \geq c\|\theta_1 - \theta_2\|_{\Theta}$ for every θ_1 and $\theta_2 \in \Theta_{\delta}$. Therefore, it follows

$$(c - o(1))\|\theta - \theta_0\|_{\Theta} \leq \|Z(\theta, h_0) - Z(\theta_0, h_0)\|_{\mathcal{L}}, \quad (13)$$

and

$$(c - o_{p^*}(1))\|\widehat{\theta} - \theta_0\|_{\Theta} \leq \|Z(\widehat{\theta}, h_0) - Z(\theta_0, h_0)\|_{\mathcal{L}} = \|Z(\widehat{\theta}, h_0)\|_{\mathcal{L}}, \quad (14)$$

with probability tending to one. By the triangle inequality and conditions **G.1** and **G.6**, the right hand side of the previous inequality is bounded by

$$\|Z(\widehat{\theta}, h_0) - Z(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} + \|Z(\widehat{\theta}, \widehat{h}) - Z(\widehat{\theta}, \widehat{h}) + Z(\theta_0, h_0) - Z(\theta_0, h_0)\|_{\mathcal{L}} + O_p(n^{-1/2}). \quad (15)$$

For the first term, we have that

$$\begin{aligned} \|Z(\widehat{\theta}, h_0) - Z(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} &\leq \|Z(\widehat{\theta}, h_0) - Z(\widehat{\theta}, \widehat{h}) - Z_2(\widehat{\theta}, h_0)[\widehat{h} - h_0]\|_{\mathcal{L}} \\ &\quad + \|Z_2(\widehat{\theta}, h_0)[\widehat{h} - h_0] - Z_2(\theta_0, h_0)[\widehat{h} - h_0]\|_{\mathcal{L}} + \|Z_2(\theta_0, h_0)[\widehat{h} - h_0]\|_{\mathcal{L}} \\ &\leq o_{p^*}(n^{-1/2}) + o_{p^*}(\|\widehat{\theta} - \theta_0\|_{\Theta}) + O_{p^*}(n^{-1/2}) \\ &\leq \|Z(\widehat{\theta}, h_0)\|_{\mathcal{L}} \times o_{p^*}(1) + O_{p^*}(n^{-1/2}), \end{aligned}$$

where the first inequality follows from the triangle inequality, the second one by conditions **G.3** and **G.6**, and the third by inequality (13).

As for the second term in (15), by condition **G.5**,

$$\begin{aligned} \|\mathbb{Z}(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}) + \mathbb{Z}(\theta_0, h_0) - \mathbb{Z}(\theta_0, h_0)\|_{\mathcal{L}} &= o_{p^*}(1/\sqrt{n}) + \|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} + \|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} \\ &= o_{p^*}(1/\sqrt{n}) + o_{p^*}(\|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}}). \end{aligned}$$

The second equality follows from condition **G.1**, $\|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} = o_{p^*}(1/\sqrt{n})$. By the triangle inequality,

$$\|\mathbb{Z}(\widehat{\theta}, \widehat{h})\|_{\mathcal{L}} \leq \|\mathbb{Z}(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}) + \mathbb{Z}(\theta_0, h_0) - \mathbb{Z}(\theta_0, h_0)\|_{\mathcal{L}} + O_{p^*}(1/\sqrt{n}).$$

It then follows

$$(1 - o_{p^*}(1))\|\mathbb{Z}(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}) + \mathbb{Z}(\theta_0, h_0) - \mathbb{Z}(\theta_0, h_0)\|_{\mathcal{L}} \leq o_{p^*}(1/\sqrt{n}).$$

Thus, equation (15) is bounded by

$$\|\mathbb{Z}(\widehat{\theta}, h_0)\|_{\mathcal{L}} \times o_{p^*}(1) + O_{p^*}(n^{-1/2}),$$

and the right side of the equality in (14) satisfies

$$(1 - o_{p^*}(1))\|\mathbb{Z}(\widehat{\theta}, h_0)\|_{\mathcal{L}} \leq O_{p^*}(n^{-1/2}). \quad (16)$$

Therefore, $(c - o_p(1))\sqrt{n}\|\widehat{\theta} - \theta_0\|_{\Theta} \leq O_{p^*}(1)$ and $\widehat{\theta}$ is \sqrt{n} -consistent for θ_0 in Θ .

Step 2: Weak Convergence

Now we show the weak convergence. By conditions **G.2** and **G.3**,

$$\begin{aligned} & \| -\mathbb{Z}(\widehat{\theta}, \widehat{h}) + \mathbb{Z}(\theta_0, h_0) - \mathbb{Z}_1(\theta_0, h_0)(\widehat{\theta} - \theta_0) - \mathbb{Z}_2(\theta_0, h_0)[\widehat{h} - h_0] \|_{\mathcal{L}} \\ &= \| -\mathbb{Z}(\widehat{\theta}, \widehat{h}) + \mathbb{Z}(\widehat{\theta}, h_0) - \mathbb{Z}_2(\widehat{\theta}, h_0)[\widehat{h} - h_0] + \mathbb{Z}(\widehat{\theta}, h_0) - \mathbb{Z}(\theta_0, h_0) - \mathbb{Z}_1(\theta_0, h_0)(\widehat{\theta} - \theta_0) \\ &\quad + \mathbb{Z}_2(\widehat{\theta}, h_0)[\widehat{h} - h_0] - \mathbb{Z}_2(\theta_0, h_0)[\widehat{h} - h_0] \|_{\mathcal{L}} \\ &\leq \| -\mathbb{Z}(\widehat{\theta}, \widehat{h}) + \mathbb{Z}(\widehat{\theta}, h_0) - \mathbb{Z}_2(\widehat{\theta}, h_0)[\widehat{h} - h_0] \|_{\mathcal{L}} + \| \mathbb{Z}(\widehat{\theta}, h_0) - \mathbb{Z}(\theta_0, h_0) - \mathbb{Z}_1(\theta_0, h_0)(\widehat{\theta} - \theta_0) \|_{\mathcal{L}} \\ &\quad + \| \mathbb{Z}_2(\widehat{\theta}, h_0)[\widehat{h} - h_0] - \mathbb{Z}_2(\theta_0, h_0)[\widehat{h} - h_0] \|_{\mathcal{L}} \\ &= o_{p^*}(n^{-1/2}) + o_{p^*}(n^{-1/2}) + o_{p^*}(n^{-1/2}) = o_{p^*}(n^{-1/2}). \end{aligned}$$

Therefore, it follows that

$$\begin{aligned}
Z_1(\theta_0, h_0)\sqrt{n}(\hat{\theta} - \theta_0) + \sqrt{n}Z_2(\theta_0, h_0)[\hat{h} - h_0] &= \sqrt{n}(-Z(\hat{\theta}, \hat{h}) + Z(\theta_0, h_0)) + o_{p^*}(1) \\
&= \sqrt{n}(\mathbb{Z}(\hat{\theta}, \hat{h}) - \mathbb{Z}(\theta_0, h_0)) + o_{p^*}(1) \\
&= \sqrt{n}(Z(\theta_0, h_0) - \mathbb{Z}(\theta_0, h_0)) + o_{p^*}(1),
\end{aligned}$$

and

$$Z_1(\theta_0, h_0)\sqrt{n}(\hat{\theta} - \theta_0) = -\sqrt{n}(Z_2(\theta_0, h_0)[\hat{h} - h_0] + (\mathbb{Z} - Z)(\theta_0, h_0)) + o_{p^*}(1) \rightsquigarrow \mathbb{G},$$

by condition **G.6**.

Now by condition **G.2** and the continuous mapping theorem, we have that

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightsquigarrow Z_1^{-1}(\theta_0, h_0)\mathbb{G}.$$

□

7.4 The Validity of the Bootstrap

A formal justification for the simulation method discussed for the two-step estimator is stated in in the main text. In the following Lemma 6 we provide a result for the validity of the bootstrap for general Z-estimator. It is also an extension of that in Chen, Linton, and Van Keilegom (2003).

There are two potential difficulties when constructing the confidence bands for the QTE. First, closed-form expressions of the covariance kernel are hard to calculate. This mainly is due to the estimation of the nuisance parameters. Second, even if closed-form expressions of the covariance kernel are available, they are useful only when the set \mathcal{T} is finite. Thus, we use the ordinary nonparametric bootstrap method to determine the rejection regions of the tests for the case when $Z(\theta, h) = \mathbb{E}m^\dagger(W_i, \theta; h(W_i, \theta))$ and $\mathbb{Z}(\theta, h) = \frac{1}{n} \sum_{i=1}^n m^\dagger(W_i, \theta; h(W_i, \theta))$, where $\{W_i\}$ is i.i.d and $m^\dagger(\cdot)$ is some known function. It is without loss of generality to study only the validity of bootstrap for $\sqrt{n}(\hat{\theta}(t) - \theta_0(t))$. Let \hat{h}^* be an estimator of h_0 using resampled data. Let $\mathbb{Z}^*(\theta, h)$ denote the resampled average. The bootstrap estimator $\hat{\theta}^*$ satisfies

$$\|\mathbb{Z}^*(\hat{\theta}^*, \hat{h}^*)\| = o_{p^*}(n^{-1/2}).$$

Following Chen, Linton, and Van Keilegom (2003), an asterisk denotes a probability or moment computed under the bootstrap distribution conditional on the original data set.

Consider the following conditions:

G.4B With P^* -probability tending to one, $\widehat{h}^* \in \mathcal{H}$ and $\|\widehat{h}^* - \widehat{h}\|_{\Pi} = o_{P^*}(n^{-1/4})$.

G.5B For any $\delta_n \downarrow 0$,

$$\sup_{\|\theta - \theta_0\| \leq \delta_n, \|h - h_0\|_{\Pi} \leq \delta_n} \|\sqrt{n}(\mathbb{Z}^* - \mathbb{Z})(\theta, h) - \sqrt{n}(\mathbb{Z}^* - \mathbb{Z})(\theta_0, h_0)\|_{\mathcal{L}} = o_{P^*}(1).$$

G.6B $\sqrt{n}(Z_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] + (\mathbb{Z}^* - \mathbb{Z})(\widehat{\theta}, \widehat{h}))$ converges weakly to a tight random element \mathbb{G} in \mathcal{L} in P^* -probability.

Conditions **G.4B–G.6B** are the bootstrap analog to the conditions to establish weak convergence.

Lemma 6. *Suppose $\theta_0 \in \text{int}(\Theta)$ and $\widehat{\theta} \xrightarrow{\text{a.s.}} \theta_0$. Assume that conditions **G.1, G.4, G.5**, and **G.6** are satisfied with “in probability” replaced by “almost surely”. Let conditions **G.2** and **G.3** hold with h_0 replaced by $h \in \mathcal{H}_{\delta_n}$. Also, assume that $Z_1(\theta; h)$ is continuous in h at $\theta = \theta_0$ and $h = h_0$. Then, under conditions **G.4B–G.6B**, $\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) \rightsquigarrow_{Z_1^{-1}(\theta_0, h_0)} \mathbb{G}$ in P^* -probability.*

Proof. The assertion that $\|\widehat{\theta}^* - \widehat{\theta}\| = O_{P^*}(n^{-1/2})$ a.s. P can be shown in a similar way as the proof of the \sqrt{n} -consistency of $\widehat{\theta}$. Therefore we omit the proof and only show the weak convergence in probability of the bootstrap estimator.

Note that

$$\begin{aligned}
& \|\mathbb{Z}^*(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}^*(\widehat{\theta}, \widehat{h}) - \mathbb{Z}_1(\widehat{\theta}, \widehat{h})(\widehat{\theta}^* - \widehat{\theta}) - \mathbb{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}]\| \\
&= \|\mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}(\widehat{\theta}^*, \widehat{h}) - \mathbb{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] + \mathbb{Z}(\widehat{\theta}^*, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}) - \mathbb{Z}_1(\widehat{\theta}, \widehat{h})(\widehat{\theta}^* - \widehat{\theta}) \\
&\quad + [(\mathbb{Z}^*(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*)) - (\mathbb{Z}^*(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}))] \\
&\quad + [(\mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}(\widehat{\theta}^*, \widehat{h})) - (\mathbb{Z}(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}))] \\
&\quad + \mathbb{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] - \mathbb{Z}_2(\widehat{\theta}^*, \widehat{h})[\widehat{h}^* - \widehat{h}]\| \\
&\leq \|\mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}(\widehat{\theta}^*, \widehat{h}) - \mathbb{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}]\| + \|\mathbb{Z}(\widehat{\theta}^*, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}) - \mathbb{Z}_1(\widehat{\theta}, \widehat{h})(\widehat{\theta}^* - \widehat{\theta})\| \\
&\quad + \|(\mathbb{Z}^*(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*)) - (\mathbb{Z}^*(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}))\| \\
&\quad + \|(\mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}(\widehat{\theta}^*, \widehat{h})) - (\mathbb{Z}(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}))\| \\
&\quad + \|\mathbb{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] - \mathbb{Z}_2(\widehat{\theta}^*, \widehat{h})[\widehat{h}^* - \widehat{h}]\| \\
&= o_{p^*}(n^{-1/2}).
\end{aligned}$$

The first term is $o_{p^*}(n^{-1/2})$ by condition **G.3** (version of this lemma) and **G.4B**. The second term is $o_{p^*}(n^{-1/2})$ by condition **G.2** (version of this lemma) and \sqrt{n} -consistency of $\widehat{\theta}^*$. The third and fourth terms are $o_{p^*}(n^{-1/2})$ by the triangle inequality and conditions **G.5'** (almost sure version) and **G.5B**. And the fifth term is $o_{p^*}(n^{-1/2})$ by condition **G.3** (version of this lemma) and \sqrt{n} -consistency of $\widehat{\theta}^*$.

Therefore, it follows

$$\begin{aligned}
\mathbb{Z}_1(\widehat{\theta}, \widehat{h})\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) + \sqrt{n}\mathbb{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] &= \sqrt{n}(\mathbb{Z}^*(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}^*(\widehat{\theta}, \widehat{h})) + o_{p^*}(1) \\
&= -\sqrt{n}(\mathbb{Z}^*(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h})) + o_{p^*}(1)
\end{aligned}$$

and

$$\mathbb{Z}_1(\widehat{\theta}, \widehat{h})\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) = -\sqrt{n}\mathbb{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] - \sqrt{n}(\mathbb{Z}^*(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h})) + o_{p^*}(1) \rightsquigarrow \mathbb{G}$$

in \mathcal{L} in P^* -probability by condition **G.6**. We can replace $\mathbb{Z}_1(\widehat{\theta}, \widehat{h})$ by $\mathbb{Z}_1(\theta_0, h_0)$ with probability one. Now by condition **G.2** (version of this lemma) and the continuous mapping theorem, we have

$$\sqrt{n}(\widehat{\theta}^* - \widehat{\theta}) \rightsquigarrow \mathbb{Z}_1^{-1}(\theta_0, h_0)\mathbb{G},$$

and the result follows. \square

8 Monte Carlo

In this section we conduct numerical experiments to evaluate the finite sample properties of the proposed methods. We report results for the empirical size and power of the uniform tests. We are mainly interested in studying the properties of the tests based on QTE over \mathcal{T} .

8.1 Experiment Design

In the experiments, we use the same data generating process (DGP) as in Firpo (2007). The generated data follow a very simple specification. Starting with $X = [X_1, X_2]^\top$, we set

$$\begin{aligned} X_1 &\sim \left[\mu_{X_1} - \frac{\sqrt{12}}{2}, \mu_{X_1} + \frac{\sqrt{12}}{2} \right] \\ X_2 &\sim \left[\mu_{X_2} - \frac{\sqrt{12}}{2}, \mu_{X_2} + \frac{\sqrt{12}}{2} \right] \end{aligned}$$

which will be independent random variables with the following means and variances: $E[X_1] = \mu_{X_1}$, $E[X_2] = \mu_{X_2}$, and $V[X_1] = V[X_2] = 1$. The treatment indicator is set to be $T = 1\{\delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_1^2 + \eta > 0\}$, where η has a logistic c.d.f. as $F(u) = (1 + \exp(-\pi u/10\sqrt{3}))^{-1}$. The potential outcomes are $Y(0) = \gamma_1 X_1 + \gamma_2 X_2 + \epsilon_0$ and $Y(1) = Y(0) + \epsilon_1 - \epsilon_0$, where ϵ_0 and ϵ_1 are, respectively, distributed as $N(0, \sigma_{\epsilon_0}^2)$ and $N(\beta, \sigma_{\epsilon_1}^2)$. The variables X , η , ϵ_0 , and ϵ_1 are mutually independent. Under this specification, $Y(1)$ and $Y(0)$ will be distributed as the sum of two uniforms and a normal.

The parameters were chosen to be $\mu_{X_1} = 1$, $\mu_{X_2} = 5$, $\delta_0 = -1$, $\delta_1 = 5$, $\delta_2 = -5$, $\delta_3 = -0.05$, $\gamma_1 = -5$, $\gamma_2 = 1$. For the simple experiment, the parameters β , $\sigma_{\epsilon_0}^2$, $\sigma_{\epsilon_1}^2$ control the testing procedure under the null and alternative.

To investigate the empirical size, we consider the above DGP with $\beta = 0$ and $\sigma_{\epsilon_0}^2 = \sigma_{\epsilon_1}^2 = 5$. To evaluate the power of the test we use two different configurations: (i) varying the parameter $\beta \in \{0, 6\}$; (ii) varying the parameter $\sigma_{\epsilon_0}^2 \in \{5, 20\}$, while keeping $\sigma_{\epsilon_1}^2 = 5$. In the later case, by using a $\sigma_{\epsilon_0}^2$ different from $\sigma_{\epsilon_1}^2$ we are able to achieve a positive treatment across the quantiles.

We implement tests for the null hypothesis that the treatment effect is ineffective. Thus, we estimate $\Delta(\tau) = q_1(\tau) - q_0(\tau)$ and test whether $\Delta(\tau) = 0$ for all τ . We report results for the Cramér-von Mises test for the simulations. The results for the Kolmogorov-Smirnov

	B=250			B=500		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$n = 500$	0.008	0.035	0.079	0.010	0.039	0.080
$n = 750$	0.009	0.045	0.084	0.010	0.046	0.085
$n = 1000$	0.009	0.048	0.091	0.011	0.049	0.092

Table 6: Size of the uniform tests ($\beta = 0$)

tests are similar. For the estimation of w_t^0 in the first-step, we use a nonparametric estimation with a local linear logit and a leave-one-out for choice of the number of polynomials.

We examine the empirical rejection frequencies for 1%, 5%, and 10% ($\alpha = \{0.01, 0.05, 0.10\}$) nominal levels tests for different choices of sample size $n = \{500, 750, 1000\}$. We also investigate different numbers of bootstraps $\{250, 500\}$. The number of replications is 2,000.

8.2 Results

We present the empirical size and power for the proposed tests. Table 6 collects the results for empirical size and Figures 3 and 4 display the empirical power functions when varying β and $\sigma_{\epsilon_0}^2$, respectively.

In Table 6 we report the empirical sizes for different samples and nominal sizes. First, we observe that the empirical sizes ($\beta = 0$) are close to the respective nominal ones, 1%, 5%, and 10%. We also study the impact of sample size and number of bootstraps on the size. The size improves with the sample size, but it is not very sensitive to the number of bootstraps, implying that smaller number of bootstraps is satisfactory. Overall, Table 6 shows that the uniform tests have good size property even in small samples.

The empirical power functions are displayed in Figures 3 and 4. In Figure 3 we vary β . The results show that the power of test improves as the sample size increases. The main point is that as the parameter β increases, the treatment increases, and so does the probability of the test rejecting the null of effect of the treatment. Figure 4 displays the results for empirical power when varying $\sigma_{\epsilon_0}^2$. The results are qualitatively similar and show that the power increases as the heterogeneity increases. In addition, as the sample size increases the improves the raises. As in the previous case, the results suggest that the number of bootstraps does not have a substantial effect on the power.

Overall the simulations show the usefulness of our uniform inference procedures in de-

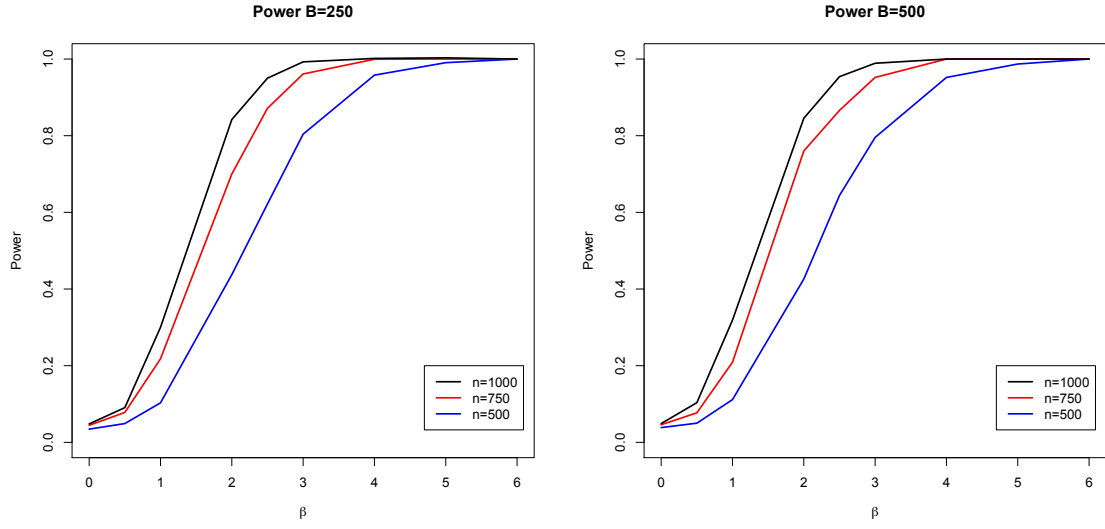


Figure 3: Empirical power function when varying β . Left box plots the power function for different sample sizes and number of bootstraps 250. Right box plots the power function for different sample sizes and number of bootstraps 500.

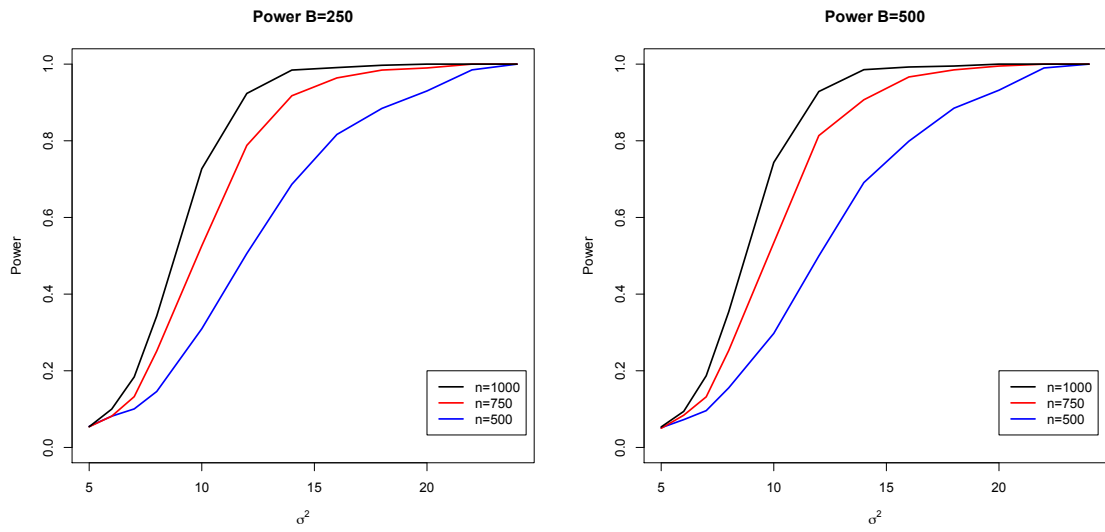


Figure 4: Empirical power function when varying $\sigma_{\epsilon_0}^2$. Left box plots the power function for different sample sizes and number of bootstraps 250. Right box plots the power function for different sample sizes and number of bootstraps 500.

tecting cases where heterogeneity is an important concern. The results suggest the proposed methods have good finite sample performance, leading to reliable, powerful, and computationally attractive inference. Our main proposal, the uniform tests, in addition to having good power properties, makes the bootstrap method a practical inference procedure.