Quantile continuous treatment effects

Javier Alejo

CEDLAS-UNLP-CONICET, Calle 6 entre 47 y 48, 5to. piso, oficina 516, B1900 La Plata, Buenos Aires, Argentina

Antonio F. Galvao

University of Arizona, Department of Economics, McClelland Hall, Room 401 1130 E. Helen Street, Tucson, AZ 85721, United States

Gabriel Montes-Rojas

CONICET-IIEP-Universidad de Buenos Aires, Av. Córdoba 2122, C1120AAQ Ciudad Autónoma de Buenos Aires, Argentina

Abstract

Continuous treatments (e.g., doses) arise often in practice. Methods for estimation and inference for quantile treatment effects models with a continuous treatment are proposed. Identification of the parameters of interest, the dose-response functions and the quantile treatment effects, is achieved under the assumption that selection to treatment is based on observable characteristics. An easy to implement semiparametric two-step estimator, where the first step is based on a flexible Box-Cox model is proposed. Uniform consistency and weak convergence of this estimator are established. Practical statistical inference procedures are developed using bootstrap. Monte Carlo simulations show that the proposed methods have good finite sample properties. Finally, the proposed methods are applied to a survey of Massachusetts lottery winners to estimate the unconditional quantile effects of the prize amount, as a proxy of non-labor income changes, on subsequent labor earnings from U.S. Social Security records. The empirical results reveal strong heterogeneity across unconditional quantiles. The study suggests that there is a threshold value in non-labor income that is high enough to make all individuals stop working, and that this applies uniformly for all quantiles. It also shows that the threshold value is monotonic in the quantiles.

Keywords: Continuous treatment, quantile treatment effects, quantile regression

Email addresses: javieralejo@gmail.com (Javier Alejo), agalvao@email.arizona.edu (Antonio F. Galvao), gabriel.montes@fce.uba.ar (Gabriel Montes-Rojas)

Corresponding author: Antonio F. Galvao. Stata computer programs to replicate the numerical analyses are available as a supplementary material.

Preprint submitted to Elsevier *October 17, 2017*

1. Introduction

The effect of policy variables on distributional outcomes are of fundamental interest in empirical economics and they are of importance for policymakers. The treatment effects (TE) literature has been extensively used in economics to analyze how treatments or social programs affect selected outcomes of interest. While the original focus was on average treatment effects (ATE), it has been largely recognized that heterogeneity in the TE is of utmost importance for policy evaluation. That is, programs need not only to be evaluated on whether they have an average impact, but on who and how beneficiaries are affected by it. This heterogeneity has been mostly analyzed in terms of quantile treatment effects (QTE) (see, e.g., Abadie et al., 2002; Chernozhukov and Hansen, 2005; Firpo, 2007; Cattaneo, 2010), a valuable method of statistical analysis where the main interest is the effect on the entire outcome distributions.

Recently, there has been a growing interest on continuous TE (CTE). In CTE models programs can be evaluated not only by treatment indicator(s) but also on the quality or length of the treatment(s). Continuous treatments (such as those indexed by dose, exposure, duration, or frequency) arise very often in practice, especially in observational studies. Importantly, such treatments lead to effects that are naturally described by curves (e.g., dose-response curves as functionals of the treatment dose) rather than scalars (e.g., point estimators) as in discrete treatments. Many papers in the literature on unconditional TE concentrate on discrete treatments, i.e. binary or multi-valued treatment assignments. On the binary TE models, Hahn (1998), Heckman et al. (1998), Hirano et al. (2003), Abadie and Imbens (2006), Imbens et al. (2006), and Li et al. (2009) study efficient estimation of the ATE. There is also literature on estimation of QTE for multi-valued TE, see e.g., Imbens (2000), Lechner (2001), Cattaneo (2010) and Cattaneo et al. (2013). It is known that categorizing or discretizing continuous treatments generally leads to a number of serious problems as loss of power in testing, misclassification (which is associated with potential bias), problems for prediction, and even interpretation of the results and coefficients of interest. See, e.g., Cox

(1957), Cohen (1983), van Belle (2008), and Fedorov et al. (2009) for more comprehensive discussions on problems associated with discretizing continuous variables. Among others, Hirano and Imbens (2004) and Imai and van Dyk (2004) develop a generalized propensity score (GPS) for continuous average treatment models. Flores (2007) proposes nonparametric estimators for average dose-response functions (ADRF). Florens et al. (2008) consider identification of ATE using control functions. Flores et al. (2012) estimate causal effects of different lengths of exposure to academic and vocational instruction. d'Haultfoeuille et al. (2015) discuss identification with CTE in nonseparable models when the causal variable of interest is endogenous. Kennedy et al. (2015) develop nonparametric methods for doubly robust estimation of continuous treatment effects.

This paper contributes to this literature by developing practical estimation and inference for QTE with a continuous treatment. A parameter of interest in the presence of continuous treatment is the entire curve of quantile potential outcomes or quantile dose-response function (QDRF). The QDRF is defined as the quantile function over the entire set of the continuous treatment values, and it summarizes the potential responses of each dose of magnitude $t \in \mathcal{T}$ on a specified outcome of interest at the unconditional quantile $\tau \in (0,1)$. Another parameter of interest is the quantile continuous treatment effect (QCTE), which corresponds, for any fixed quantile, to the difference between two QDRF's at given levels of treatment. That is, the QTE at a given dose level and for a corresponding magnitude (extra dose). In this paper, we focus on estimating the QDRF and QCTE. Recent work on which we build our analysis has focused on estimating these functionals. Galvao and Wang (2015) provide a general statistical framework to evaluate general functionals based on continuous TE using Z-estimators; Lee (2015) studies the unconditional distribution of potential outcomes with continuous treatments using the partial mean processes, and this can be applied to quantile operators.

Identification of the parameters of interest is based on the ignorability or weak unconfoundedness assumption applying the methodology of Galvao and Wang (2015). Following

Rosenbaum and Rubin (1983), the relevant restriction for identification is the ignorability assumption, that is, the selection to treatment is based on observable variables (i.e., pre-treatment covariates related to treatment assignment and outcome). The ignorability assumption states that given a set of observed covariates, each individual is randomly assigned to either the treatment group or the control group. This condition has been largely employed in the literature, see, e.g., Rubin (1977), Heckman et al. (1998), Dehejia and Wahba (1999), Firpo (2007), and Flores (2007). The resulting identification equation depends on the moment condition given by the influence function of the quantile estimation multiplied by a ratio of conditional density functions. The use of this ratio is related to Hirano and Imbens (2004) and Lee (2015) approach to construct a GPS to allow for a continuous treatment.

The empirical estimators are constructed based on the identification equation and are implemented as two-step estimators. The practical implementation of the estimator is simple. In the first step, one estimates a ratio of conditional densities. In the second step of the two-step estimator, a simple weighted quantile regression estimation is performed where the weights are given by the ratio of conditional density functions. We derive the asymptotic properties of the two-step estimator, namely, uniform consistency and weak convergence. Since the treatment is continuous and the treatment levels take values on an interval $\mathcal{T}$, we establish the results uniformly over the set of treatment values. Based on these asymptotic results we develop statistical inference procedures for uniform inference and for fixed treatment values of interest. These can be implemented using bootstrap methods, for which we provide explicit algorithms.

Different alternatives have been explored in the literature for estimating the GPS and conditional densities in general for models based on the ignorability assumption. Galvao and Wang (2015) suggest a nonparametric estimation for the first step. However, there are issues with its practical implementation. First, these procedures require to properly estimate densities for "many" covariates, as the ignorability assumption is usually only valid after a rich set of control variables are considered. Nonparametric density estimators are computationally difficult for high-dimensional settings, and are thus problematic to implement in practice. Second, the required rates of convergence of the nonparametric estimator might be difficult to achieve. Overcoming these difficulties Flores et al. (2012) propose a parametric but flexible estimation of the GPS based on generalized linear models. Their proposed estimator relies on selecting an appropriate model to estimate the density for the GPS. In this paper, we follow Wei and Carroll (2009) and use an alternative estimation for the first step. In particular, we implement a flexible Box-Cox density estimation procedure. This follows Wand et al. (1991) where a transformation of the variables before density estimation is performed, so that this estimate of the density of the transformed variable is "back-transformed" to an estimate of the density of the original variable. The Box-Cox approach has important advantages. First, the Box-Cox first step is simple to implement in practice. Second, the Box-Cox procedure allows for many covariates and satisfies the required converge rates for the first step. The Box-Cox is thus very flexible to accommodate empirical settings where the ignorability assumption is only valid after conditioning on a rich (possibly large) set of covariates.

Monte Carlo simulations evaluate the finite sample performance of the methods for estimating QDRF and QCTE curves. The numerical simulations show numerical evidence of good performance. The estimators are approximately unbiased and consistent. The evidence clearly shows that the Box-Cox procedure is a flexible procedure to correctly estimate QDRF and QCTE functions for alternative data generating processes.

To illustrate the methods we estimate the effects of non-labor income changes on labor earnings. We use the data from Imbens et al. (2001) who study the effect of lottery prizes on labor market variables. They use the survey of Massachusetts lottery winners and estimate the effect of the prize amount, as a proxy of exogenous non-labor income changes, on subsequent labor earnings (from U.S. Social Security records). This database has also been used in Hirano and Imbens (2004), Bia and Mattei (2008) and Bia et al. (2014) for estimating ADRF. Their results show that non-labor income monotonically reduces future labor earnings. The lottery prize, being unrelated with labor market performance, conditional on a rich set of observables, serves as an income shock that may be used to measure the income effect on labor market decisions. In this example we have interest in identifying the effect of the lottery prize, which is a continuous variable, on labor earnings, and as such in estimating the QDRF and QCTE curves. That is, rather than studying the effect on a treatment group (i.e. with income shock) with respect to a comparable control group, we are interested in the curve linking labor market variables with the size of the shock. We focus on yearly income size years after the prize was received. By analyzing the quantile process we show important heterogeneity in the marginal effects of the lottery prize. In particular, higher quantiles of future labor market earnings are less responsive to an increment in the lottery prize than lower quantiles. Our analysis also reveals the prize threshold value that makes the individual stop participating in the labor market is monotonic on the quantiles. These results are important for analyzing the effect of general income transfers, as conditional cash transfer programs in developing countries, as the quantile heterogeneity reveals that those that are more likely to opt out of the labor market are the ones in the lower part of the income distribution.

The remainder of the paper is organized as follows. Section 2 presents the model and establishes the conditions for identification. Section 3 discusses the details of the construction of the proposed two-step estimator. Section 4 derives the asymptotic properties of the two-step estimator. Inference procedures using the bootstrap are discussed in Section 5. Monte Carlo simulations are provided in Section 6. The empirical application appears in Section 7, and Section 8 concludes. All the proofs are collected in the Appendix.

2. The model and identification

The model's objective is to learn how an outcome variable changes as the dose of some treatment variable varies. The dose is denoted by t, where $t \in \mathcal{T}$, an interval in $\mathbb{R}$, and the outcome is denoted by $Y(t)$. More specifically, for each $t \in \mathcal{T}$, $Y(t)$ is the outcome when the dose of treatment is t. Thus define the random process $Y(t)$ as t varies in $\mathcal{T}$. In the binary treatment case $\mathcal{T} = \{0, 1\}$. Here we allow $\mathcal{T}$ to be an interval $[t_0, t_1]$.

An important parameter of interest when the treatment is continuous is the quantile dose response function (QDRF), which is defined as

$$q_\tau(t) \in \inf\{q : F_{Y(t)}(q) \geq \tau\}, \quad \tau \in (0, 1), \tag{1}$$

the unconditional τ-th QDRF, where $F_{Y(t)}$ is the distribution function of $Y(t)$. Thus, the QDRF summarizes the potential responses of each dose of magnitude $t \in \mathcal{T}$ on a specified outcome of interest, $Y(t)$, at its unconditional quantile τ.

From the QDRF, one can learn about another interesting parameter, the quantile continuous treatment effect (QCTE), which is defined as

$$\Delta_\tau(t, t') := q_\tau(t) - q_\tau(t'). \tag{2}$$

The QCTE, as defined in (2), captures the difference of the τ-th quantile at two given different levels of treatment, t and t'. This QCTE function is the same as defined in Lee (2015) and describes the difference between the two potential responses of $Y(t)$ at doses of magnitude t and t', at a given unconditional quantile τ. Note that, in this paper, the QCTE is defined as the difference of the τ-th quantile at different levels of treatment. This definition does not require the assumption of rank preservation, and it is regarded as a convenient way to summarize interesting aspects of marginal distributions of the potential outcomes. However, if rank preservation holds, then QCTE defined above has a causal interpretation, that is, the effect of changing the level of the treatment for any particular subpopulation. We refer the reader to Firpo (2007) and Cattaneo (2010) for a detailed discussion on rank preservation in quantile treatment effects and definitions of concepts. Of particular interest is to analyze the QCTE for a fixed change in the dose, say δ, over the doses $t \in \mathcal{T}$ as

$$D_\tau(t, \delta) := \Delta_\tau(t + \delta, t) = q_\tau(t + \delta) - q_\tau(t). \tag{3}$$

One could approximate the quantile partial effect, $\frac{\partial}{\partial t} q_\tau(t)$, by a difference quotient, with $\Delta_\tau(t, t')$ divided by $(t - t')$. As noted by an anonymous referee another parameter of interest is $\Delta_\tau(t, t_0)$ for a fixed value of t_0. If we use $t_0 = 0$ then this would estimate the effect of receiving t amount of dose (as compared to no dose).

Unfortunately, as usual in the treatment effects literature, one cannot observe $Y(t)$ for all $t \in \mathcal{T}$. Rather, only a single $Y(t_0)$ can be observed, where t_0 is the realization of a random variable T. Hence, if assignment to treatment status depends on potential outcomes, as it is usual in economic and other non-experimental problems, then selection biases arises as the observed outcomes might not be the result of the dose itself but of a self-assignment into treatment.

To solve this problem, it is common in the TE literature to assume the existence of a set of random variables $\boldsymbol{X}$ conditional on which $Y(t)$ is independent from T for all $t \in \mathcal{T}$. Thus conditional on observable variables, observed outcomes can be given a causal interpretation. This is the ignorability condition or weak unconfoundedness assumption in the literature. Finally, we need to combine the results for $\boldsymbol{X}$ to obtain an unconditional TE. By the law of iterated expectations, unconditional expectations can be recovered. This is summarized in the following assumption:

Assumption I.1 For all $t \in \mathcal{T}$, $Y(t) \perp T | \boldsymbol{X}$.

According to Assumption **I.1**, although the assignment of the treatment level is not random, it is random within subpopulations characterized by $\boldsymbol{X}$. This assumption has been extensively employed in the literature, among others, by Heckman et al. (1998), Dehejia and Wahba (1999), and Hirano and Imbens (2004).

The continuous TE model differs from binary or multivalued TE models. In our case we require the conditional density function of the treatment, a continuous variable, conditional on observables $(\boldsymbol{X}, Y) \in (\mathcal{X}, \mathcal{Y})$, to be positive:

Assumption I.2 For all $t \in \mathcal{T}$, $f_{T|\boldsymbol{X},Y}(t|\boldsymbol{x}, y) > 0$ for $\boldsymbol{x} \in \mathcal{X}$ and $y \in \mathcal{Y}$.

Assumption **I.2** requires the conditional density of the treatment, conditional on the observed outcome and the covariates, to have positive mass over the relevant region. This is slightly different from the usual assumption in the TE literature that uses the generalized propensity score in which only $f_{T|\boldsymbol{X}}(t|\boldsymbol{x}) > 0$ is imposed. In our case, to establish identification and replace the unobserved $Y(t)$ by the observed Y, we require the density of the treatment conditional on observables to have positive mass.

Define $m(Y(t); q_\tau(t)) = \tau - \mathbf{1}\{Y(t) < q_\tau(t)\}$ for each t and let

$$\mathrm{E}[m(Y(t); q_\tau(t))] = 0.$$

The identification result is presented in the following lemma. The result is a direct application of the Theorem 1 in Galvao and Wang (2015) who extended the propensity score method to general dose response functions in a setting with continuous treatment. The intuition behind the result is that $Y(t)$ being unobserved is replaced with observables $(\boldsymbol{X}, Y, T)$ equipped with a proper estimation of the density function of the treatment conditional on $(\boldsymbol{X}, Y)$.

The next lemma states the identification of the QDRF, $q_\tau(t)$.

Lemma 1 (Identification). *Under assumptions* **I.1**–**I.2**, *and assuming that*

(i) For each $t \in \mathcal{T}$, $q_{\tau 0}(t)$ *uniquely solves* $\mathrm{E}[m(Y(t); q_\tau(t))] = 0$, *where* $m : \mathbb{R}^2 \mapsto \mathbb{R}$ *is measurable;*

(ii) There exists a function $e(y)$ *with* $\int e(y)\, dy < \infty$ *such that* $|m(y; q_\tau(t_0)) f_{T,Y|\boldsymbol{X}}(t_0 + \delta t, y|\boldsymbol{x})| \le e(y)$ *and* $\mathrm{E}[m(Y; q_\tau(t_0))|\boldsymbol{X}, T = t_0] = \lim_{\delta t \downarrow 0} \mathrm{E}[m(Y; q_\tau(t_0))|\boldsymbol{X}, T \in [t_0, t_0 + \delta t]]$. *Also the interval* $\mathcal{T}$ *is right open.*

We have

$$\mathrm{E}[m(Y(t); q_\tau(t))] = \mathrm{E}\left[m(Y; q_\tau(t)) w_0(\boldsymbol{U}; t)\right] \tag{4}$$

for each $t \in \mathcal{T}$, *where* $w_0(\boldsymbol{u}; t) := \frac{f_{T|\boldsymbol{X},Y}(t|\boldsymbol{x},y)}{f_{T|\boldsymbol{X}}(t|\boldsymbol{x})}$ *and for notational convenience we denote* $\boldsymbol{u} := (\boldsymbol{x}^\top, y)^\top$ *and* $\boldsymbol{U} := (\boldsymbol{X}^\top, Y)^\top$. *Consequently,*

$$\mathrm{E}\left[m(Y; q_\tau(t)) w_0(\boldsymbol{U}; t)\right] = 0 \tag{5}$$

if and only if $q_\tau(t) = q_{\tau 0}(t)$.

Condition (i) in Lemma 1 is an identification condition which is common in the QR literature. Condition (ii) allows for changing the orders of limits and integral. The set $\mathcal{T}$ is right open without loss of generality. The result in equation (4) allows identification of the QDRF. The left hand side of (4) is used to define $q_\tau(t)$, which involves the unobservable $Y(t)$. Consequently, it cannot be used to estimate $q_\tau(t)$. Nevertheless, the right hand side of (4) is expressed in terms of the observables $(\boldsymbol{X}, Y, T)$, and hence, can be used to estimate $q_\tau(t)$. Note that $Y(t)$ is not observable while Y is. The intuition behind the result is that the existence of $\boldsymbol{X}$ delivers identification of the parameter of interest. That is, conditional on observed covariates $\boldsymbol{X}$, each individual is randomly assigned to a treatment level.

As in the TE literature, the identification induces an estimating equation with two pieces, the function $m(\cdot)$ together with a weighting function $w_0(\cdot)$. In our case, the weights are given by $\frac{f_{T|\boldsymbol{X},Y}(t|\boldsymbol{x},y)}{f_{T|\boldsymbol{X}}(t|\boldsymbol{x})}$. The intuition of this result is similar to the discrete case where the propensity score is replaced by the corresponding density function. Also note that the weights could be written as $\frac{f_{Y|\boldsymbol{X},T}(y|\boldsymbol{x},t)}{f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})}$. In either case, we need to work with a ratio of two conditional densities. Note that this approach seems different from Hirano and Imbens (2004) and other papers that followed, where they only estimate $f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$, the so called generalized propensity score. However, Hirano and Imbens approach also requires to estimate $E[Y|X,T]$, or in fact, $E[Y|f_{T|X}(t|x),T]$. As such, ours and Hirano and Imbens' procedures involve two different functional estimates to compute the parameter of interest.

Finally, since the QCTE is the difference between the QDRF at two different treatment doses, identification of QCTE, $\Delta_\tau(t,t')$, is a straightforward consequence of Lemma 1, as stated in the next corollary.

Corollary 1. *Identification of Quantile Continuous Treatment Effect Parameters: Under the assumptions of Theorem 1, the quantile continuous treatment effect parameter* $\Delta_{\tau 0}(t,t')$ *for any* $(t,t') \in \mathcal{T}^2$ *are identified from data on* $(\mathcal{X}, \mathcal{Y}, \mathcal{T})$.

3. Two-step estimators

Using the identification expression from Lemma 1, we are able to construct two-step estimators for both QDRF and QCTE, in equations (2) and (3) respectively, as in Firpo (2007), Cattaneo (2010) and Galvao and Wang (2015), by estimating equation (5). The estimation method is a two-step procedure. In the first step one estimates the weights, that is, the ratio of densities, $w(\boldsymbol{u};t) := \frac{f_{T|\boldsymbol{X},Y}(t|\boldsymbol{x},y)}{f_{T|\boldsymbol{X}}(t|\boldsymbol{x})}$. The second step is given by a reweighed version of the standard quantile estimation procedure (Koenker and Bassett (1978)). Below we describe the details of estimation.

We have a random sample of units $(\boldsymbol{X}_i, Y_i, T_i)$, indexed by $i = 1, ..., n$. For each unit i, $\boldsymbol{X}_i$ is a vector of covariates, and the level of the treatment received is $T_i \in [t_0, t_1]$. We observe the vector $\boldsymbol{X}_i$, the treatment received T_i, and the observed outcome corresponding to the level of the treatment received, Y_i.

3.1. First step: Estimation of w_0

To implement the estimator in practice we need an estimator for w_0, i.e., $\widehat{w}(\boldsymbol{u}, t) = \frac{\widehat{f}_{T|X,Y}(t|x,y)}{\widehat{f}_{T|X}(t|x)}$. It is common in the literature to estimate nuisance parameters in two-step estimators using parametric models in the first step, see e.g., Murphy and Topel (1985), Newey and McFadden (1994), Chernozhukov and Hong (2002), Hirano and Imbens (2004), Wei and Carroll (2009), Montes-Rojas (2009) and Flores et al. (2012). We follow this literature and also propose a flexible parametric approach for the first step estimation.

In practice, one estimates $f_{T|\boldsymbol{X},y}(t|\boldsymbol{x}, y)$ and $f_{T|\boldsymbol{X}}(t|\boldsymbol{x})$ separately, and then computes the ratio to estimate w_0. Galvao and Wang (2015) suggest a potential nonparametric estimation for the first-step. However, there are important issues with its practical implementation. First, in most empirical applications the number of variables in $\boldsymbol{X}$ used to satisfy the ignorability condition is relatively large, and as it is well known in the literature, the dimension of $\boldsymbol{X}$ has an adverse effect on nonparametric methods due to the curse of dimensionality. Hence, practical estimation might be infeasible. Second, as the conditions we state below will show, when using nonparametric estimation in the first-step, the convergence rate of the estimator might be slower than required. In addition, in particular treatment effects examples, a formal verification of the required conditions might be very difficult when dealing with nonparametric estimation. Therefore, there are compelling reasons to use flexible parametric models to estimate the ratio of the conditional density functions. In this subsection we suggest a flexible Box-Cox estimation.

Wand et al. (1991) argue that kernel density estimators do not perform well when the underlying density has features that require different amounts of smoothing at different locations. Following the results of Carroll and Ruppert (1984), they propose a transformation of the variables before density estimation is performed, so that this estimate of the density of the transformed variable is "back-transformed" to an estimate of the density of the original variable. The Box-Cox transformations of both sides of a regression model has been successfully implemented in many contexts ever since (see Wang and Ruppert (1995) and Wei and Carroll (2009)). The Box-Cox approach has important advantages. First, the Box-Cox first step is simple to implement in practice. Second, the Box-Cox procedure allows for many covariates and satisfies the required convergence rates for the first step.

To estimate the conditional density $f_{T|\boldsymbol{X},Y}(t|\boldsymbol{x}, y)$, we use the following model

$$\Lambda(T, \lambda_1) = \Lambda((\boldsymbol{X}), \lambda_2)\beta_X + \Lambda(Y, \lambda_2)\beta_Y + \epsilon, \tag{6}$$

where $\epsilon|\boldsymbol{X}, Y \sim N(0, \sigma_\epsilon^2)$, and $\Lambda(\cdot, \lambda)$ is the Box-Cox transformation function, which is defined as $\Lambda(Z, \lambda) = \log Z$ if $\lambda = 0$ and $= \frac{Z^\lambda - 1}{\lambda}$ otherwise. Using maximum likelihood estimation, we obtain the unknown parameters $\mu := (\lambda_1, \lambda_2, \boldsymbol{\beta}_X, \beta_Y, \sigma_\epsilon^2)$, and finally the conditional density, $\widehat{f}_{T|\boldsymbol{X},Y}(t|\boldsymbol{x}, y)$, from the normality assumption. Similarly, one estimates $f_{T|\boldsymbol{X}}(t|\boldsymbol{x})$. We refer the reader to Wei and Carroll (2009, p.1133) for details of the density estimation using Box-Cox procedures when the transformed model is assumed Gaussian.

The Box-Cox transformation only applies to variables in a positive domain (excluding zero). Nevertheless, this could be implemented if we define, for a given variable x, $x^* = e^x$, where we could thus have negative, zero and positive values of x, and we allow the Box-Cox parameters to transform x^*. In this case, if the estimated parameter λ is indeed zero, then the variable would require no transformation. Our Monte Carlo simulations show that the Box-Cox Gaussian model performs well for a large family of distributions. It is important to highlight that the normality assumption is a simplifying condition. Nevertheless one can employ any other distribution of the Box-Cox estimation. That is, we could implement a likelihood estimation of the Λ functions above imposing alternative distribution of ϵ. Another point of potential concern is the fact that the estimator may not be well behaved when the true $\lambda = 0$. Our Monte Carlo experiments, however, perform very well in this particular case.

3.2. Second step: Estimation of $q_{\tau 0}$ and $\Delta_{\tau 0}$

The resulting identification condition for $q_{\tau 0}(t)$ from Theorem 1 is: $\mathrm{E}[(\tau - \mathbf{1}\{Y < q_{\tau 0}(t)\})w_0(\boldsymbol{U};t)] = 0$. Thus, an estimator for the QDRF $q_{\tau 0}(t)$ is

$$\widehat{q}_\tau(t) = \arg\min_q \frac{1}{n}\sum_{i=1}^{n} \widehat{w}\left(\boldsymbol{u}_i;t\right)\rho_\tau\left(y_i - q\right), \tag{7}$$

where $\rho_\tau(\cdot) := \cdot(\tau - \mathbf{1}\{\cdot < 0\})$ is the check function as in Koenker and Bassett (1978). Practical implementation of the estimator is simple. In practice, given the random sample, $(\boldsymbol{X}, T, Y)$, one first computes $\widehat{w}$ in the first step as described previously. Then, in the second step, one computes a simple weighted quantile regression of Y on a constant term using $\widehat{w}$ as weights as given in equation (7), for each given t taken over a discretized subset (i.e., grid) of $\mathcal{T}$.

Estimation of the QCTE parameter, $\Delta_{\tau 0}(t,t')$, is also easy. Given the QDRF $\widehat{q}_\tau(t)$, the estimator $\widehat{\Delta}_\tau(t,t')$ can be computed as

$$\widehat{\Delta}_\tau(t,t') = \widehat{q}_\tau(t) - \widehat{q}_\tau(t'), \tag{8}$$

for any $(t,t') \in \mathcal{T}^2$.

4. Asymptotic properties of the two-step estimators

In this section, we derive the asymptotic properties of the two-step estimators. Since the treatment is continuous and the treatment levels take values on an interval $\mathcal{T}$, we establish the uniform consistency and the weak limit of the QDRF, $\widehat{q}_\tau(\cdot)$, in $\ell^\infty(\mathcal{T})$. Given this result, consistency of the QCTE, $\widehat{\Delta}_\tau(\cdot,\cdot)$ follows. In addition, we discuss the practical estimation of the nuisance parameter, w_0. **Notations:** Let E and $\mathbb{E}_n$ 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 $|g(x)|_\infty$ denote $\sup_x |g(x)|$ for $x \in \mathcal{X} \subset \mathbb{R}^d$, where d is a positive integer.

4.1. Conditions for the estimator of w_0

As discussed in the previous section, estimation of w_0 in the first step is important for practical implementation. It is also important for inference that $\widehat{w}$ satisfies the theoretical conditions discussed above. For estimators of w_0 to have the desirable properties, we impose the following general assumptions.

A.1 Let the model in equation (6) hold,

A.2 $\boldsymbol{X} \in \mathbb{R}^d$, $(\boldsymbol{X}, Y) \in (\mathcal{X}, \mathcal{Y})$, and $\sigma_\epsilon^2 < \infty$.

Condition **A.1** imposes the Box-Cox model as a valid approximation to the weights, and **A.2** only requires finite σ_ϵ^2. Conditions **A.1** and **A.2** imply several properties of the estimated weights necessary to establish consistency and asymptotic normality of the two-step estimator. The following proposition summarizes the result.

Proposition 1. *Under* **A.1** *and* **A.2***, the following properties hold:*

PC.1 *The function class $\{\psi_{w,t} : w \in w_\delta, t \in \mathcal{T}\}$, where $\psi_{w,t} = w(\boldsymbol{U}; t)$, is Glivenko-Cantelli, and has an envelope that is integrable.*

PC.2 *There exists $0 < M_w < \infty$ such that $w_0(\boldsymbol{u}; t) < M_w$.*

PC.3 $|\widehat{w} - w_0|_\infty = o_p(1)$.

PG.1 *The function class* $\{\psi_{w,t} : w \in w_\delta, t \in \mathcal{T}\}$ *is Donsker with a uniform bound.*

PG.2 $\sqrt{n}\mathrm{E}[(\tau - \mathbf{1}\{Y < q_{\tau 0}(t)\})(w(\boldsymbol{U}; t) - w_0(\boldsymbol{U}; t))]|_{w=\widehat{w}}$ *converges weakly.*

PG.3 $|\widehat{w} - w_0|_\infty = o_p(n^{-1/4})$.

Results **PC.1–PC.3** and **PG.1–PG.3** are statistical properties of the nuisance parameter. Properties **PC.1–PC.3** are used to establish consistency of the second step, and **PG.1–PG.3** are important to derive weak convergence. These results show that the Box-Cox estimator considered in (6) belongs to the class of functions usually used for first-stage estimation in the semiparametric literature. Given the implications from Proposition 1, in the next subsection we state the results using conditions **PC.1–PC.3** and **PG.1–PG.3** directly.

The estimators proposed on Section 3.1 are recommended for practical use. In particular, the Box-Cox two-side transformation of the regression model of $T|\boldsymbol{X}$ and $T|\boldsymbol{X}, Y$, density estimation and back-transformation satisfies the required conditions.

4.2. Consistency and weak convergence

Consider first consistency of the QDRF uniformly over $t \in \mathcal{T}$. Consider the following condition.

PC.4 Uniformly in t, the densities $f_{Y(t)}(y)$ is bounded above and $f_{Y(t)}(q_{\tau 0}(t)) > 0$. Also, for any $\delta > 0$, there exists ϵ_δ such that $\inf_{|q - q_{\tau 0}|_\infty > \delta} |\mathrm{E}[(\tau - 1\{Y < q\})w_0(\boldsymbol{U}; t)]_\infty > \epsilon_\delta$.

PC.4 is a standard identification condition in the quantile regression literature. It is similar to A.2-A.3 of Angrist et al. (2006), and corresponds to Assumption 2 of Firpo (2007). Cattaneo (2010) uses a similar assumption for quantile estimation in the multivalued treatment effects setting. The next result states consistency of the two-step estimtor.

Theorem 1 (Uniform consistency). *The two-step estimator of the QDRF is uniformly consistent, i.e., as* $n \to \infty$, $|\widehat{q}_\tau(\cdot) - q_{\tau 0}(\cdot)|_\infty = o_{p^*}(1)$, *provided* **PC.1**–**PC.4** *are satisfied.*

Given the result in Theorem 1, consistency of the $\widehat{\Delta}_\tau(t,t')$ is given in the following corollary.

Corollary 2. *Under Assumptions of Theorem 1, for any* $(t,t') \in \mathcal{T}^2$, *as* $n \to \infty$,

$$\widehat{\Delta}_\tau(t,t') \xrightarrow{p} \Delta_{\tau 0}(t,t').$$

Now we state the weak convergence of QDRF result.

Theorem 2 (Weak convergence). *Suppose that* $|\mathrm{E}[m(Y;q_{\tau 0}(t))w_0(\boldsymbol{U};t)]|_\infty = 0$, $|\widehat{q}_\tau - q_{\tau 0}|_\infty = o_{p^*}(1)$, *and that conditions,* **PG.1–PG.3** *are satisfied. Then, in* $\ell^\infty(\mathcal{T})$, *as* $n \to \infty$,

$$\sqrt{n}(\widehat{q}_\tau(t) - q_{\tau 0}(t)) \rightsquigarrow \mathbb{G}(t),$$

where $\mathbb{G}(t)$ *is a mean zero Gaussian process with covariance function*

$$\mathrm{E}[\mathbb{G}(t)\mathbb{G}(t')'] = \mathrm{Z}_1^{-1}(t)S(t,t')\mathrm{Z}_1^{-1}(t'),$$

with $\mathrm{Z}_1(q_0,w_0)$ *being the derivative of* $q \mapsto \mathrm{E}[m(Y;q)w_0(\boldsymbol{U};\cdot)]$ *at* q_0, *i.e.* $\frac{\partial \mathrm{E}[m(Y,q(t))w_0(\boldsymbol{U};t)]}{\partial q(t)}|_{q(t)=q_0(t)}$, *and*

$$\begin{aligned} S(t,t') \;=\; & \mathrm{E}\left[(w_0(\boldsymbol{U};t)\left(m(Y,q_0(t)) - \mathrm{E}\left[m(Y,q_0(t))|X,T=t\right]\right) + \mathrm{E}\left[m(Y,q_0(t))|X,T=t\right]\right) \\ & \cdot \left(w_0(\boldsymbol{U};t')\left(m(Y,q_0(t')) - \mathrm{E}\left[m(Y,q_0(t'))|X,T=t'\right]\right) + \mathrm{E}\left[m(Y,q_0(t'))|X,T=t'\right]\right)\right]. \end{aligned}$$

Finally, the asymptotic normality of the $\widehat{\Delta}_\tau(t,t')$ follows from Theorem 2 and is given in the corollary.

Corollary 3. *Under Assumptions of Theorem 2, for any* $(t,t') \in \mathcal{T}^2$, *as* $n \to \infty$,

$$\sqrt{n}\left(\widehat{\Delta}_\tau(t,t') - \Delta_{\tau 0}(t,t')\right) \rightsquigarrow \mathbb{G}_\Delta(t,t')$$

where $\mathbb{G}_\Delta$ *is a mean zero Gaussian process with covariance function*

$$\begin{aligned} \mathrm{E}\left[\mathbb{G}_\Delta(t)\mathbb{G}_\Delta(t')'\right] \;=\; & \mathrm{Z}_1^{-1}(t)S(t,t)[\mathrm{Z}_1^{-1}(t)]' - \mathrm{Z}_1^{-1}(t)S(t,t')[\mathrm{Z}_1^{-1}(t')]' \\ & - \mathrm{Z}_1^{-1}(t')S(t',t)[\mathrm{Z}_1^{-1}(t)]' + \mathrm{Z}_1^{-1}(t')S(t',t')[\mathrm{Z}_1^{-1}(t')]'. \end{aligned}$$

5. Inference procedures

In this section, we turn our attention to inference on both the QDRF and QCTE. 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. First, we consider inference for point estimation for a fixed level of treatment. Second, uniform inference over the set of treatments $\mathcal{T}$.

5.1. Inference for fixed t

First, we consider inference procedures for the QDRF for a fixed t, where we test simple linear hypothesis as

$$H_0 : q_{\tau 0}(t) = q_0,$$

for a fixed treatment t, where q_0 is a scalar value of interest, e.g., $q_0 = 0$. Inference for these simple hypotheses can be based on the results of Theorem 2 and, in particular, on the asymptotic normality of $\sqrt{n}(\widehat{q}_\tau(t) - q_0)$.

Consider now asymptotic inference on QCTE. Since the QCTE involves evaluating two different treatment values, say t and t', simple hypothesis testing can be stated as

$$H_0 : \Delta_\tau(t, t') = \Delta_0,$$

which is based on the procedures for the QDRF estimator. The formal justification for this procedure is based on the application of the continuous mapping theorem on the results from QDRF. As noted above, a particular object of interest in the QCTE analysis is the comparison of two treatment values separated by a fixed interval, $\Delta_\tau(t, t + \delta) = D_\tau(t, \delta)$. Note that for fixed t and t' (or $t + \delta$), inference procedures can be seen as an extension of Firpo (2007) inference for binary treatments in QTE models when we consider two different treatment levels.

5.1.1. Implementation of testing procedures

The practical implementation of the procedures will be based on the bootstrap. Consider B random samples from the original sample data $\{(Y_i, \boldsymbol{X}_i, T_i), i = 1, \ldots, n\}$. For each

$b = 1, \dots, B$:

Bootstrap algorithm

(i) Obtain the resampled data $\{(Y_i^b, \boldsymbol{X}_i^b, T_i^b), i = 1, \dots, n\}$.

(ii) Estimate the QDRF $\widehat{q}_\tau^b(t)$ or QCTE $\widehat{\Delta}_\tau^b(t, t+\delta)$ for all $t \in \mathcal{T}$.

Given this bootstrap procedure, simple hypotheses testing for fixed t can be based on Wald statistics. For instance, define the statistic of test as $\sqrt{n}\frac{\widehat{q}_\tau(t)-q_0}{\sqrt{\widehat{Var(\widehat{q}_\tau^B(t))}}}$ and $\sqrt{n}\frac{\widehat{\Delta}_\tau(t,t+\delta)-\Delta_0}{\sqrt{Var(\widehat{\widehat{\Delta}_\tau^B(t,t+\delta)})}}$, for QDRF and QCTE respectively. Now from the bootstrap procedure one is able to consistently estimate $\widehat{Var(\widehat{q}_\tau^B(t))} = \frac{1}{B}\sum_{b=1}^B(\widehat{q}_\tau^b(t) - \widehat{q}_\tau(t))^2$ for QDRF, and $Var(\widehat{\widehat{\Delta}_\tau^B(t, t+\delta)}) = \frac{1}{B}\sum_{b=1}^B(\widehat{\Delta}_\tau^b(t, t+\delta) - \widehat{\Delta}_\tau(t, t+\delta))^2$ for the QCTE. Finally, the critical values are given by a standard normal table. Consistency of this bootstrap procedure is given in Chen et al. (2003) and we omit the details.

5.2. Uniform inference

Since we consider continuous treatment, a more interesting inference for the QDRF is uniform on $\mathcal{T}$. First, for testing QDRF, we consider the following general null hypothesis

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

uniformly, where $r(t)$ is assumed to be known, continuous in t over $\mathcal{T}$, and $r \in \ell^\infty(\mathcal{T})$. For instance, a basic hypothesis of interest might be that the treatment impact summarized by $q_\tau(t)$ is ineffective for all doses, i.e. statistically equal to zero for all $t \in \mathcal{T}$. The alternative is that the treatment differs from zero at least for some $t \in \mathcal{T}$.

Inference is carried out uniformly over the set of treatment levels, $\mathcal{T}$. The basic inference process is

$$Q_n(t) := \widehat{q}_\tau(t) - r(t), \quad t \in \mathcal{T}.$$

General hypotheses on the vector $q_\tau(t)$ can be accommodated through functions of $Q_n(\cdot)$. We consider the Kolmogorov and Cramér-von Mises type test statistics, $T_n = f(Q_n(\cdot))$, where

$f(\cdot)$ represents the functionals for those two test statistics, as

$$T_{1n} := \sqrt{n} \sup_{t \in \mathcal{T}} |Q_n(t)|, \quad T_{2n} := \sqrt{n} \int_{t \in \mathcal{T}} |Q_n(t)| \, dt.$$

These statistics and their associated limiting theory provide a natural foundation for testing the null hypothesis. It is possible to formulate a wide variety of tests using variants of the proposed tests.

Now we present the limiting distributions of the test statistics under the null hypothesis. From Theorem 2 and under the null hypothesis ($H_0 : q_{\tau 0}(t) = r(t)$), it follows that $\sqrt{n}(\widehat{q}_\tau(t) - r(t)) \rightsquigarrow Z_1^{-1}\mathbb{G}(t) \equiv \bar{\mathbb{G}}(t)$. Thus, the following corollary summarizes the limiting distributions.

Corollary 4. *Assume the conditions of Theorem 2. Under $H_0 : q_{\tau 0}(t) = r(t)$, as $n \to \infty$,*

$$T_{1n} \rightsquigarrow \sup_{t \in \mathcal{T}} |\bar{\mathbb{G}}(t)|, \quad T_{2n} \rightsquigarrow \int_{t \in \mathcal{T}} |\bar{\mathbb{G}}(t)| \, dt.$$

For uniform inference for QCTE we consider general null hypothesis

$$H_0 : \Delta_{\tau 0}(t, t+\delta) - r(t) = 0, \quad t \in \mathcal{T},$$

uniformly, where δ is a fixed treatment increment, $r(t)$ is assumed to be known, continuous in t over $\mathcal{T}$, and $r \in \ell^\infty(\mathcal{T})$. Inference is carried uniformly over the set of treatment levels, $\mathcal{T}$. The basic inference process is

$$D_n(t) := \widehat{\Delta}_\tau(t, t+\delta) - r(t), \quad t \in \mathcal{T}.$$

As before we consider Kolmogorov and Cramér-von Mises type test statistics, $T_n = f(D_n(\cdot))$, where $f(\cdot)$ represents the functionals for those two test statistics, as

$$T_{3n} := \sqrt{n} \sup_{t \in \mathcal{T}} |D_n(t)|, \quad T_{4n} := \sqrt{n} \int_{t \in \mathcal{T}} |D_n(t)| \, dt.$$

The weak limit of T_{3n} and T_{4n} are given in the following result.

Corollary 5. *Assume the conditions of Lemma 3. Under* $H_0 : \Delta_{\tau 0}(t+\delta, t) - r(t) = 0$, $t \in \mathcal{T}$, *and a fixed* $\delta > 0$, *as* $n \to \infty$,

$$T_{3n} \rightsquigarrow \sup_{t \in \mathcal{T}} |\bar{\mathbb{G}}(t+\delta) - \bar{\mathbb{G}}(t)|, \quad T_{4n} \rightsquigarrow \int_{t \in \mathcal{T}} |\bar{\mathbb{G}}(t+\delta) - \bar{\mathbb{G}}(t)|, dt.$$

The weak limits of all the test statistics in Corollaries 4 and 5 are functionals of Gaussian processes, and the estimation of their covariance kernel is difficult in practice. Therefore, to make practical inference we suggest the use of simple bootstrap techniques to approximate the limiting distribution.

5.2.1. Implementation of testing procedures

For uniform inference, for all $\mathcal{T}$, the procedure is implemented in a discretized subset, most conveniently on intervals of equal size, $\mathcal{T} = [t_1, ..., t_m]$, $t_1 < ... < t_m$. Then consider the bootstrap algorithm described in the previous subsection for fixed t, and apply this for all $t \in \mathcal{T}$. Galvao and Wang (2015) suggest the following scheme.

Bootstrap algorithm

(i) For each $b = 1, ..., B$, compute $Q_n^b = (\widehat{q}_\tau^b(t) - \widehat{q}_\tau(t))$ or $D_n^b = (\widehat{\Delta}_\tau^b(t, t+\delta) - \widehat{\Delta}_\tau(t, t+\delta))$ for all $t \in \mathcal{T}$.

(ii) Compute the statistic of test of interest: $\widehat{T}_{1n}^b = \max_{t \in \mathcal{T}} \sqrt{n}|V_n^b|$, $\widehat{T}_{2n}^b = ave_{t \in \mathcal{T}} \sqrt{n}|V_n^b|$, $\widehat{T}_{3n}^b = \max_{t \in \mathcal{T}} \sqrt{n}|W_n^b|$ and $\widehat{T}_{4n}^b = ave_{t \in \mathcal{T}} \sqrt{n}|W_n^b|$.

(iii) Construct the critical values $\widehat{C}_{j,1-\alpha}$ as the $1-\alpha$ quantiles of $\{\widehat{T}_{jn}^b\}_{b=1}^B$, for $j = 1, 2, 3, 4$.

(iv) Reject if $\widehat{T}_{jn} > \widehat{C}_{j,1-\alpha}$, for $j = 1, 2, 3, 4$.

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

QG.IB For any $\delta_n \downarrow 0$, $\sup_{||w - w_0||_\Pi \leq \delta_n} |\frac{1}{n}\sum_{i=1}^n w(\boldsymbol{u}; t) - \mathrm{E}[w_0(\boldsymbol{u}; t)]|_\infty = o_{p^*}(1/\sqrt{n})$.

QG.IIB $\sqrt{n}\frac{1}{n}\sum_{i=1}^n [(\tau - \mathbf{1}\{Y_i < q_{\tau 0}(t)\})(\widehat{w}^*(\boldsymbol{U}_i; t) - \widehat{w}(\boldsymbol{U}_i; t))]$, where $\widehat{w}^*(\boldsymbol{U}_i; t)$ is the bootstrap estimate, converges weakly to a tight random element $\mathbb{G}$ in $\ell^\infty(\mathcal{T})$ in P^*-probability.

Proposition 2. *Under the conditions of Theorem 2 and* **QG.IB–QG.IIB** *and* **QG.3** *with "in probability" replaced by "almost surely", the bootstrap estimator of the QDRF is $\sqrt{n}$-consistent and converges weakly in $\ell^\infty(\mathcal{T})$, with $\sqrt{n}(\widehat{q}^*_\tau(t)-\widehat{q}_\tau(t)) \rightsquigarrow \mathrm{Z}_1^{-1}(q_{\tau 0}(t), w_0(\boldsymbol{U};t))\mathbb{G}(\tau)$ in P^*-probability.*

Proposition 2 establishes the consistency of the bootstrap procedure, and it is an extension of ordinary nonparametric bootstrap in Chen et al. (2003). This bootstrap procedure is similar to the nonparametric bootstrap of Rothe (2010, theorem 3) and Firpo and Pinto (2016, theorem 4), where bootstrap samples are obtained with replacement from the original data.

It is important to highlight the connection between this result and the previous section. Proposition 2 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. Thus, computation of critical values and practical inference are feasible.

6. Monte Carlo simulations

This section conducts simulations to investigate the finite sample performance of the QDRF and QCTE estimators. For comparison purposes, we also report simulations for the ADRF and ACTE estimators.

6.1. Experiment designs

Consider a simple data generating process. Let $v_{ji} \sim (0, \sigma_j^2)$, $j = 1, 2, 3$ and $i = 1, 2, ..., n$, and define

$$X_i = v_{1i} + 20,$$

$$T_i = X_i + v_{2i},$$

$$Y_i = T_i + X_i + (\gamma_1 + \gamma_2(T_i - 20)^2)v_{3i}.$$

We set $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$. For $(\gamma_1 = 1, \gamma_2 = 0)$ we have the location shift model and for $(\gamma_1 = 1/5, \gamma_2 = 1/5)$ a location-scale shift model. We consider two data generating processes (DGP). First we set the innovations $v_i \sim N(0,1), i = 1,2,3$, i.e. standard normal random variables, and second standardized chi-squared with 3 degrees of freedom $v_i \sim \frac{\chi_3^2 - 3}{\sqrt{6}}, i = 1,2,3$.

These DGP explicitly define a random variable $Y(t)|X$ where $ADRF(t) = \mathrm{E}[Y(t)] = 20 + t$ and $ACTE(t,t') = t' - t$. The QDRF and QCTE functions have a simple analytical solution for the Gaussian case in which the QDRF is

$$q_\tau(t) = 20 + t + \sqrt{\sigma_1^2 + (\gamma_1 + \gamma_2(t-20)^2)^2 \sigma_3^2}\,\Phi(\tau)^{-1},$$

and the QCTE is

$$\Delta_\tau(t,t') = (t - t') + \left(\sqrt{\sigma_1^2 + (\gamma_1 + \gamma_2(t-20)^2)^2\sigma_3^2} - \sqrt{\sigma_1^2 + (\gamma_1 + \gamma_2(t'-20)^2)^2\sigma_3^2}\right)\Phi(\tau)^{-1},$$

while for the chi-squared case it depends on the τ-quantile of $\nu_{1i} + (\gamma_1 + \gamma_2(t-20)^2)\nu_{3i}$, which are computed by simulations.

We provide two different estimators for the first step for comparison reasons. First, we estimate both $f_{T|X}(t|X)$ and $f_{T|X,Y}(t|X,Y)$ in a regression model assuming Gaussian error terms, and thus $w(t)$ corresponds to a ratio of normal density functions. In this case, we construct the estimates as

$$\widehat{f}_{T|X}(t|X) = \frac{1}{\sqrt{2\pi\widehat{\sigma}^2_{T|X}}} exp\left(-\frac{(t - (\widehat{\alpha}_0 + \widehat{\alpha}_1 X))^2}{2\widehat{\sigma}^2_{T|X}}\right),$$

$$\widehat{f}_{T|X,Y}(t|X,Y) = \frac{1}{\sqrt{2\pi\widehat{\sigma}^2_{T|X,Y}}} exp\left(-\frac{(t - (\widehat{\beta}_0 + \widehat{\beta}_1 X + \widehat{\beta}_2 Y))^2}{2\widehat{\sigma}^2_{T|X,Y}}\right),$$

where $(\widehat{\alpha}_0, \widehat{\alpha}_1)$ and $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$ are the regression coefficients of a regression of T on X and (X,Y), respectively, and $\widehat{\sigma}^2_{T|\cdot}$ is the sample variance of the conditional model of T. These will be referred as the Gaussian model in the figures and tables. Note that this model serves as a benchmark for the location shift model where the innovations correspond to standard normal random variables.

Second, to compute the estimates of the densities of interest, we implement the Box-Cox regression model for both $f_{T|X}(t|X)$ and $f_{T|X,Y}(t|X,Y)$, with different variables' transformation parameter for the dependent variable and the independent variables (see Section 3, estimation of w_0). Thus, the following estimators are implemented

$$\widehat{f}_{T|X}(t|X) = \frac{t^{\widehat{\lambda}_{1X}-1}}{\sqrt{2\pi\widehat{\sigma}^2_{T|X}}} exp\left(-\frac{(\Lambda(t,\widehat{\lambda}_{1X}) - (\widehat{\beta}_{0X} + \widehat{\beta}_{1X}\Lambda(X,\widehat{\lambda}_{2X})))^2}{2\widehat{\sigma}^2_{T|X}}\right),$$

$$\widehat{f}_{T|X,Y}(t|X,Y) = \frac{t^{\widehat{\lambda}_{1YX}-1}}{\sqrt{2\pi\widehat{\sigma}^2_{T|X,Y}}} exp\left(-\frac{(\Lambda(t,\widehat{\lambda}_{1YX}) - (\widehat{\beta}_{0YX} + \widehat{\beta}_{1YX}\Lambda(Y,\widehat{\lambda}_{2YX}) + \widehat{\beta}_{2YX}\Lambda(X,\widehat{\lambda}_{2YX})))^2}{2\widehat{\sigma}^2_{T|X,Y}}\right),$$

where $(\widehat{\lambda}_{1X}, \widehat{\lambda}_{2X}, \widehat{\beta}_{0X}, \widehat{\beta}_{1X}, \widehat{\sigma}^2_{T|X})$ and $(\widehat{\lambda}_{1YX}, \widehat{\lambda}_{2YX}, \widehat{\beta}_{0YX}, \widehat{\beta}_{1YX}, \widehat{\beta}_{2YX}, \widehat{\sigma}^2_{T|X,Y})$ are the corresponding Box-Cox parameters estimators for densities $f_{T|X}(t|X)$ and $f_{T|X,Y}(t|X,Y)$, respectively. These will be referred as the Box-Cox model in the figures and tables.

In both cases we estimate the ratio of the densities as

$$\widehat{w}(X,Y,t) = \widehat{f}_{T|X,Y}(t|X,Y)/\widehat{f}_{T|X}(t|X).$$

The second step is implemented by a weighted average for ADRF and weighted quantile for QDRF, where in both cases we use the corresponding $\widehat{w}(X,Y,t)$ as weights. Finally, ACTE and QCTE are then computed as the difference of the estimated unconditional average or quantile, respectively.

We use 500 Monte Carlo repetitions for estimating the ADRF-QDRF and ACTE-QCTE functions for $t \in \{18, 18.2, ..., 21.8, 22\}$ and for sample sizes $n \in \{100, 5000\}$. The results of experiments with $n = 500$ and $n = 1000$ are similar to the reported ones and they are available upon request. For ACTE-QCTE we consider changes of size 0.2 in t. We evaluate $\tau \in \{0.25, 0.50, 0.75\}$. For space considerations we only report experiments for the location shift model when the DGP is Gaussian, and location-scale shift model for DGP chi-squared. The other experiments are available from the authors upon request. We provide results for the average estimates together with the 95% confidence interval obtained from the simulations for each estimator of interest. The results are collected in Figures 1–8 and Tables

1–2. In the figures, the solid black line is the true value of the DRF and CTE functions, the solid grey line correspond to the average estimated value of those functions at every t, and the dashed lines the 95% confidence intervals constructed from the Monte Carlo simulations. The tables report the mean integrated squared error (MISE) calculated over the interval used for t, and this is decomposed into integrated squared bias (ISBias) and integrated variance (IVar).

6.2. Results

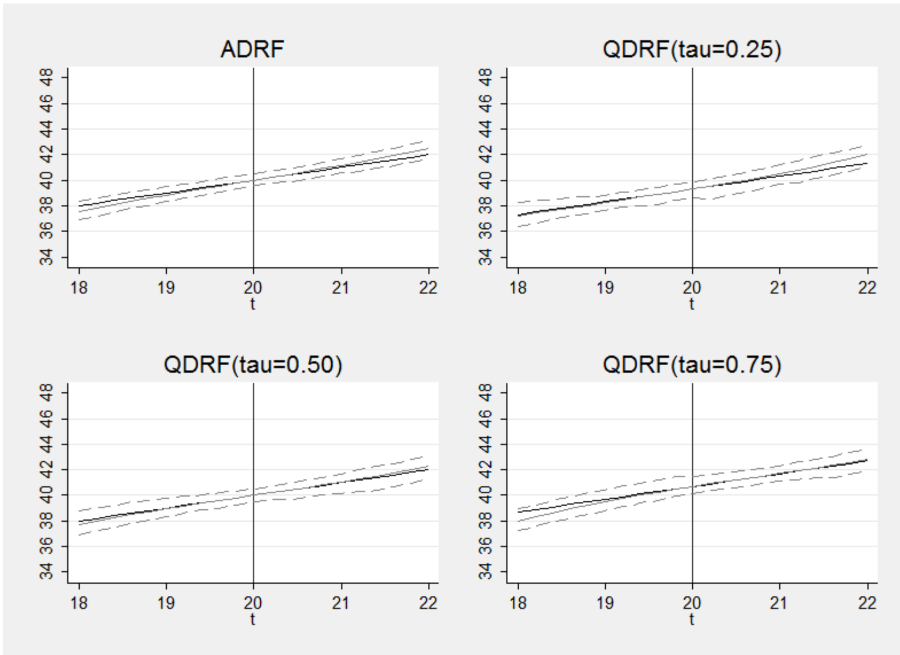
Consider first the simulations for the location shift model when the true DGP is normal. Figures 1 and 3 show results DRF for the Gaussian and Box-Cox models, respectively. These results show evidence that the estimates of the ADRF and QDRF (for all quantiles) functions are approximately unbiased. Figures 2 and 4 show results for the ACTE and QCTE functions, and they also reveal approximately unbiased estimates.

Moreover, Figures 1 to 4 show that both QDRF and QCTE models produce similar 95% intervals for both Box-Cox and Gaussian estimation methods. This result shows that the Box-Cox model is close to the Gaussian benchmark model when the true DGP is indeed Gaussian. The Figures also show that the bias is reduced as the sample size increases form $N = 100$ to $N = 5000$ for all values of t. Table 1 summarizes the ADRF and QDRF simulations calculating MISE, ISBias and IVar, for different sample sizes. The simulations clearly show that the deviations from the true value reduce as the sample size increases, either when we use the normal densities or the Box-Cox estimators in the first-step.

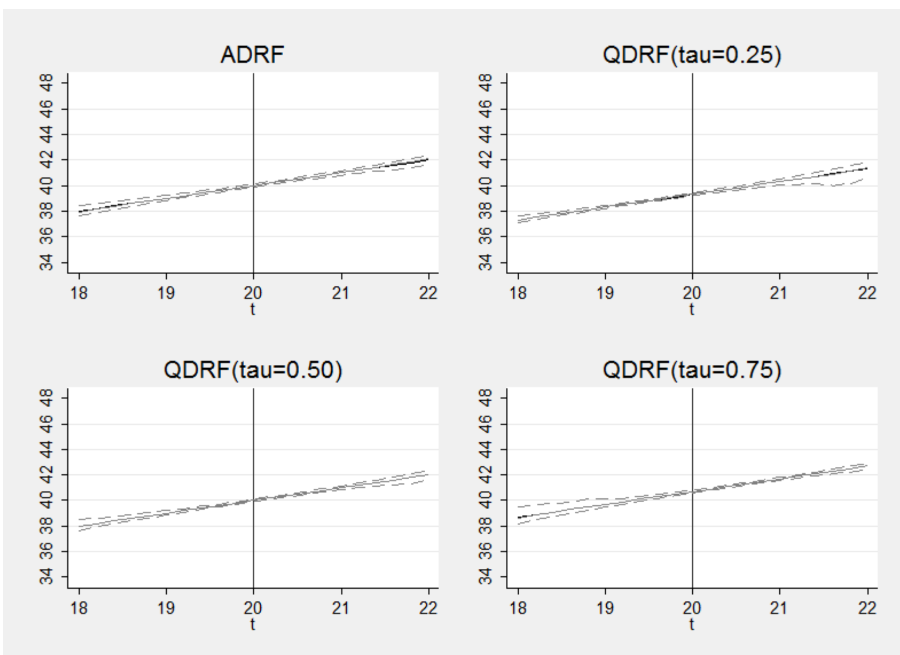
Second, we consider a location-scale shift model with chi-squared DGP. Figures 5 and 7 display the results for the DRF, and Figures 6 and 8 compile the results for CTE. In this case, there is a marked difference between the Gaussian and the Box Cox model. The figures show that assuming normal density of the error terms in the regression model results in a very large variance and significant bias of the functional estimates, although the true values lie inside the 95% confidence interval. In contrast, the Box-Cox model greatly improves the estimation, and this method is in fact able to produce unbiased estimates of the true

Figure 1: DRF: Monte Carlo simulations, location shift model, DGP: $N(0,1)$, Gaussian model

$n = 100$



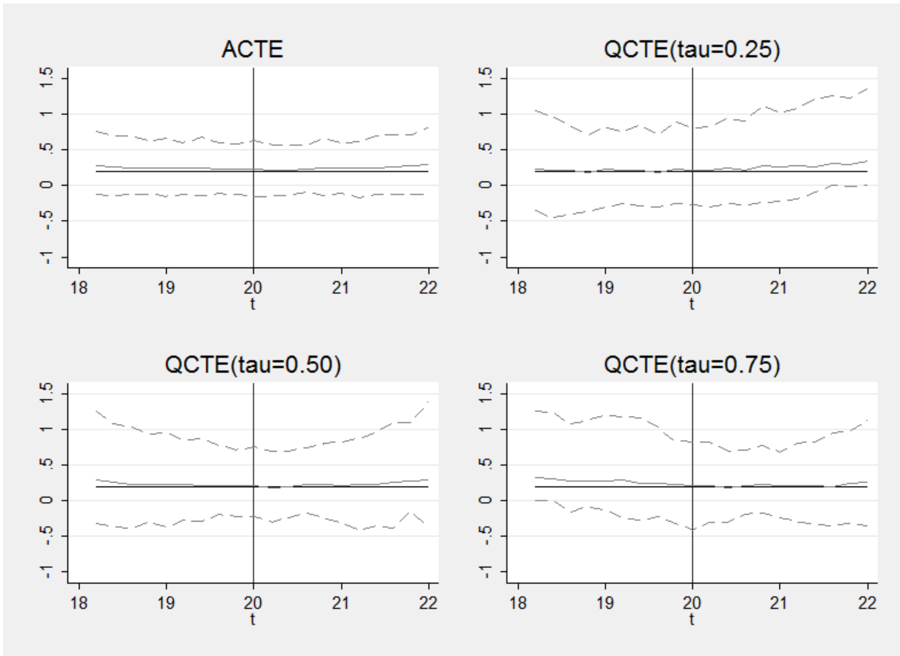
$n = 5000$



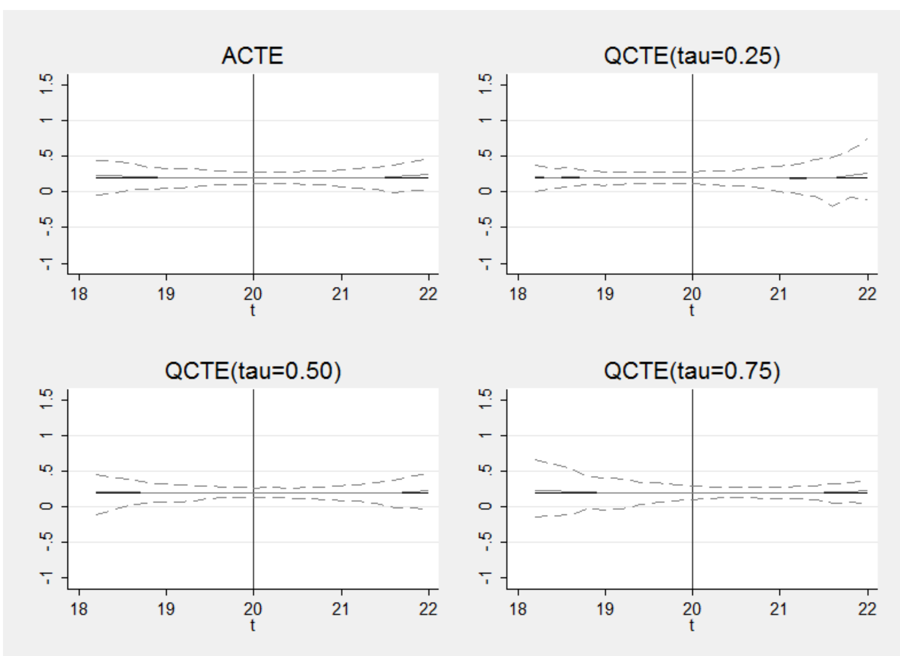
Notes: Monte Carlo experiments based on 500 simulations. In each figure the solid black line is the true value, the solid grey line correspond to the average estimated value of those functions at every t, and the dashed lines the 95% confidence intervals constructed from the Monte Carlo simulations.

Figure 2: CTE: Monte Carlo simulations, location shift model, DGP: $N(0,1)$, Gaussian model

$n = 100$



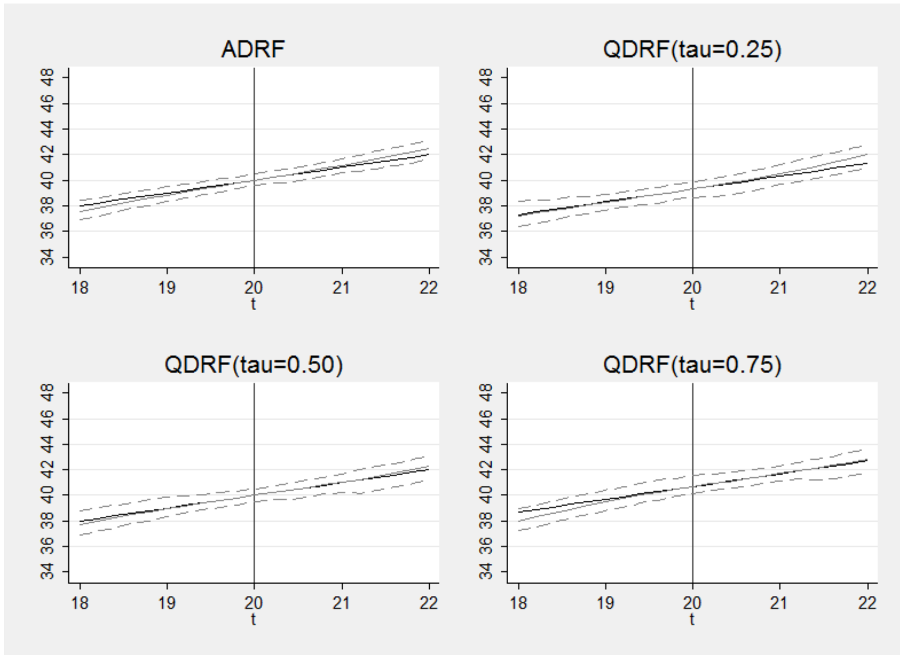
$n = 5000$



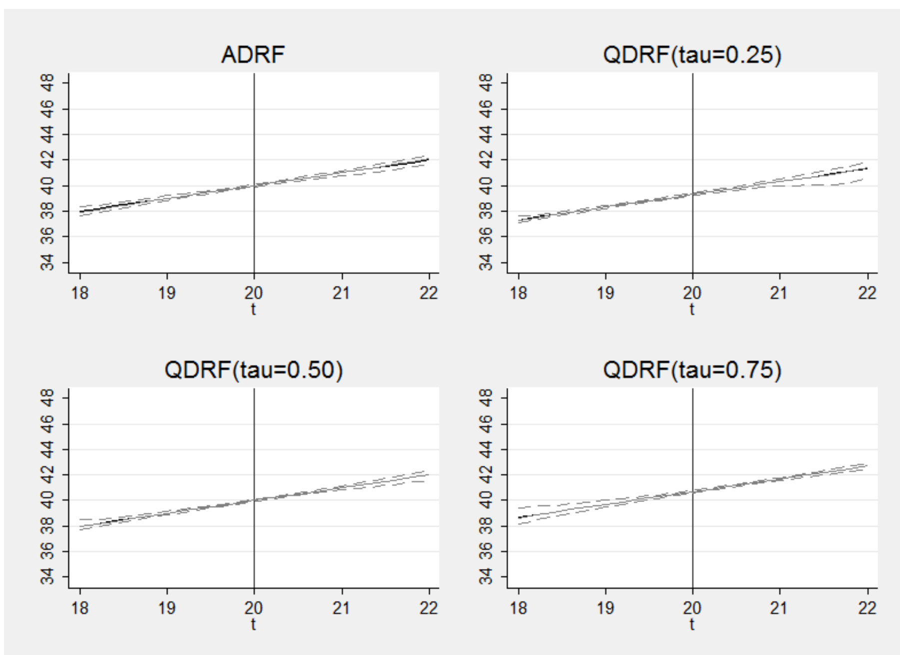
Notes: Monte Carlo experiments based on 500 simulations. In each figure the solid black line is the true value, the solid grey line correspond to the average estimated value of those functions at every t, and the dashed lines the 95% confidence intervals constructed from the Monte Carlo simulations.

Figure 3: DRF: Monte Carlo simulations, location shift model, DGP: $N(0,1)$, Box-Cox model

$n = 100$



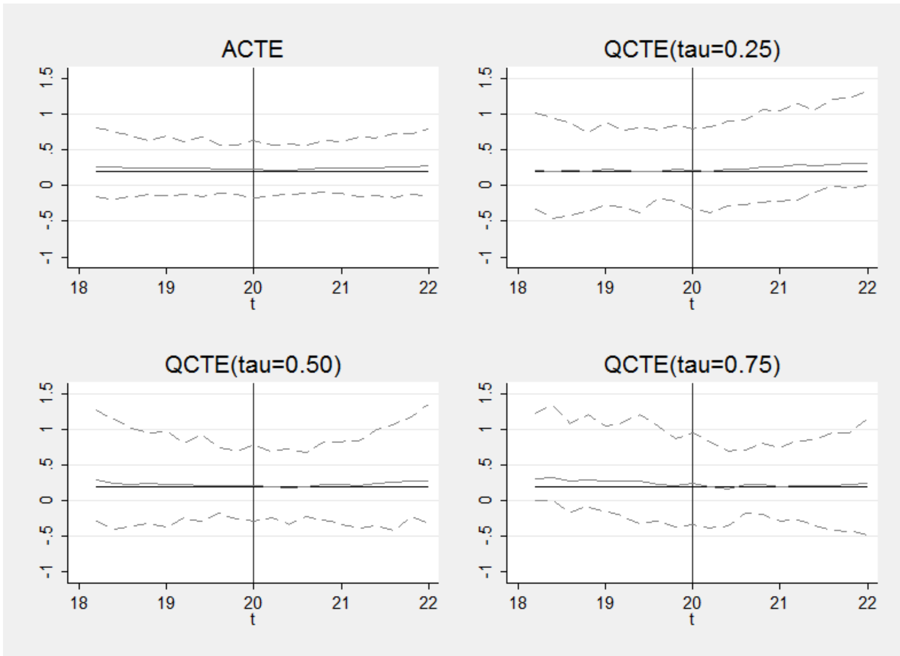
$n = 5000$



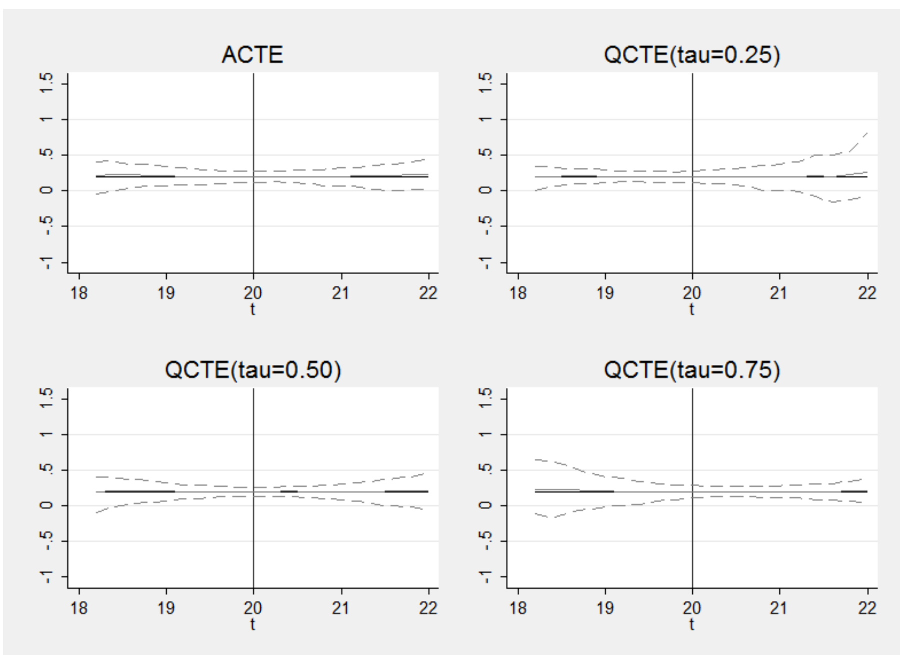
Notes: Monte Carlo experiments based on 500 simulations. In each figure the solid black line is the true value, the solid grey line correspond to the average estimated value of those functions at every t, and the dashed lines the 95% confidence intervals constructed from the Monte Carlo simulations.

Figure 4: CTE: Monte Carlo simulations, location shift model, DGP: $N(0,1)$, Box-Cox model

$n = 100$



$n = 5000$



Notes: Monte Carlo experiments based on 500 simulations. In each figure the solid black line is the true value, the solid grey line correspond to the average estimated value of those functions at every t, and the dashed lines the 95% confidence intervals constructed from the Monte Carlo simulations.

Table 1: Monte Carlo experiments, location shift model. DGP: $N(0,1)$.

n	Method	ADRF			QDRF25			QDRF50			QDRF75		
		MISE	ISBias	IVar	MISE	ISBias	IVar	MISE	ISBias	IVar	MISE	ISBias	IVar
100	Gaussian	0.201	0.058	0.143	0.302	0.062	0.239	0.255	0.019	0.236	0.290	0.061	0.229
100	Box-Cox	0.198	0.055	0.143	0.297	0.061	0.235	0.255	0.017	0.239	0.293	0.061	0.232
500	Gaussian	0.094	0.014	0.080	0.128	0.009	0.119	0.126	0.002	0.124	0.131	0.012	0.120
500	Box-Cox	0.088	0.015	0.073	0.120	0.010	0.111	0.111	0.003	0.108	0.129	0.012	0.117
1000	Gaussian	0.062	0.008	0.054	0.087	0.004	0.084	0.080	0.000	0.079	0.082	0.005	0.078
1000	Box-Cox	0.060	0.009	0.051	0.080	0.004	0.075	0.078	0.001	0.077	0.087	0.006	0.082
5000	Gaussian	0.033	0.001	0.031	0.048	0.000	0.048	0.039	0.000	0.039	0.050	0.000	0.050
5000	Box-Cox	0.029	0.002	0.027	0.049	0.000	0.049	0.034	0.000	0.034	0.037	0.001	0.037

Notes: Monte Carlo experiments based on 500 simulations.

functionals for all t. Table 2 also shows that the improvement of the Box-Cox model occurs both in terms of integrated bias and variance. Then while the Gaussian model has very large MISE values, the Box Cox reduce them to a half from $N = 100$ to $N = 5000$.

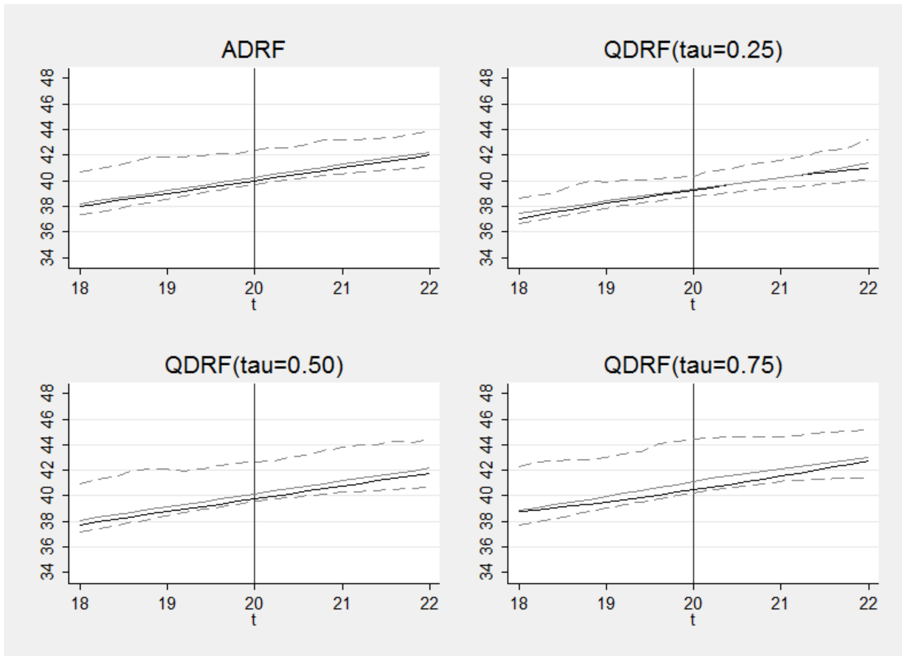
Overall, the simulations suggest that the Box-Cox model is a promising alternative to estimate density functions in the presence of covariates. It produces similar estimates to the benchmark case when the true DGP is Gaussian in the location shift model. Moreover, and more importantly, it is able to eliminate bias for other non-Gaussian distributions (in our case chi-squared) in the location-scale shift model.

7. Empirical application: the Imbens-Rubin-Sacerdote lottery sample

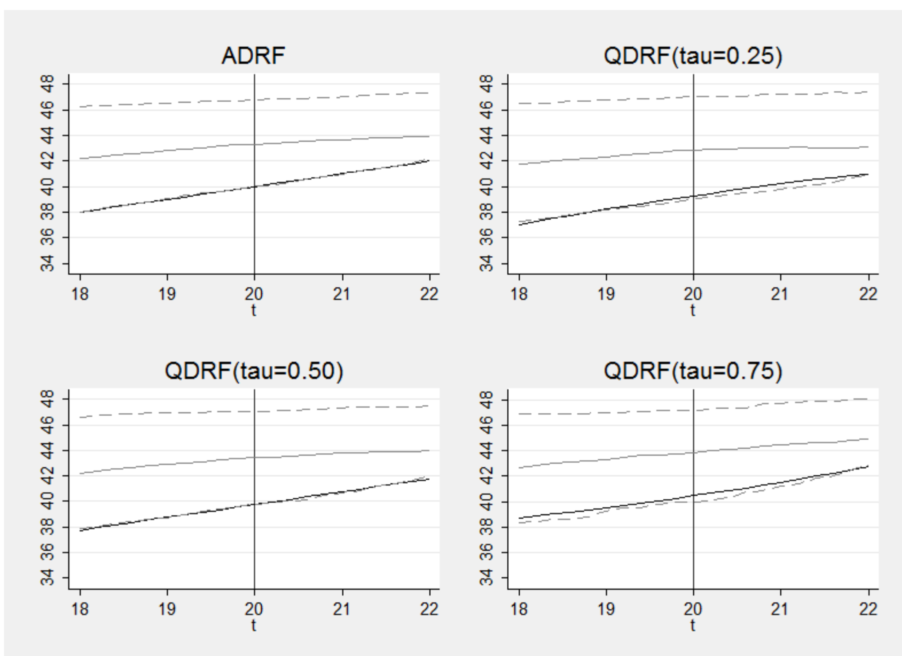
We illustrate the QCTE estimator using the survey of Massachusetts lottery winners to estimate the effect of the prize amount (as a proxy of non-labor income) on subsequent labor earnings from U.S. Social Security records. The prize amount is a continuous variable, and hence we apply the methods developed above to measure its effect on the quantiles of the distribution of earnings, also a continuous variable. This database is described in Imbens et al. (2001) and is also used as an empirical application in Hirano and Imbens (2004), Bia and Mattei (2008) and Bia et al. (2014) for estimating average dose-response functions

Figure 5: DRF: Monte Carlo simulations, location-scale shift model, DGP: $\frac{\chi_3^2-3}{\sqrt{6}}$, Gaussian model

$n = 100$



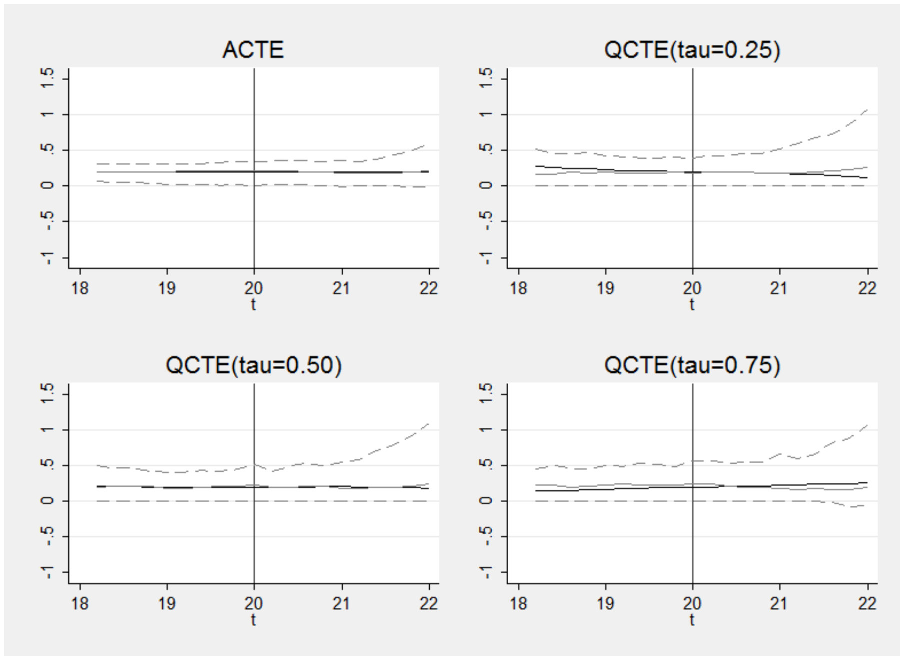
$n = 5000$



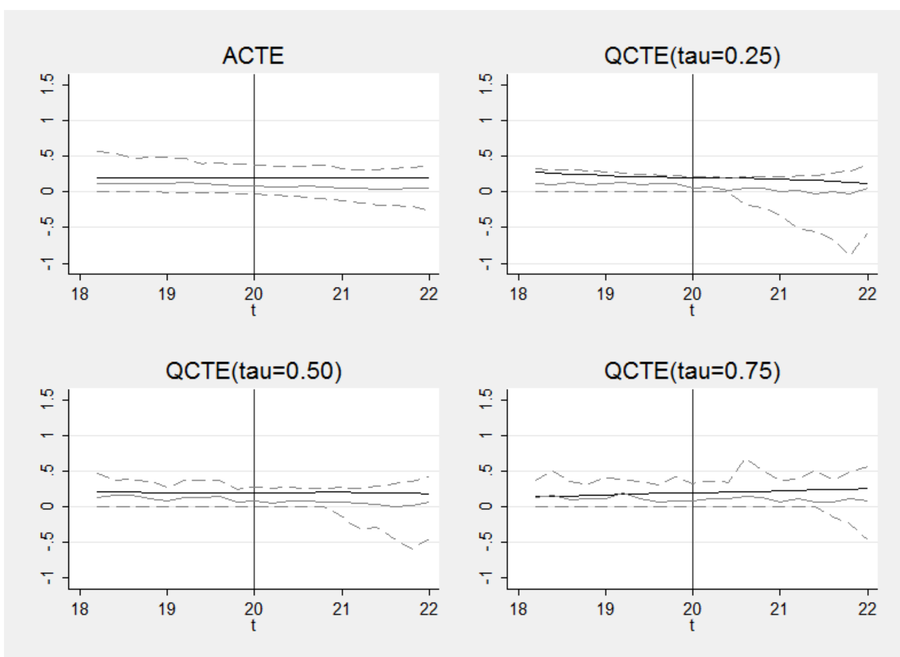
Notes: Monte Carlo experiments based on 500 simulations. In each figure the solid black line is the true value, the solid grey line correspond to the average estimated value of those functions at every t, and the dashed lines the 95% confidence intervals constructed from the Monte Carlo simulations.

Figure 6: CTE: Monte Carlo simulations, location-scale shift model, DGP: $\frac{\chi_3^2-3}{\sqrt{6}}$, Gaussian model

$n = 100$



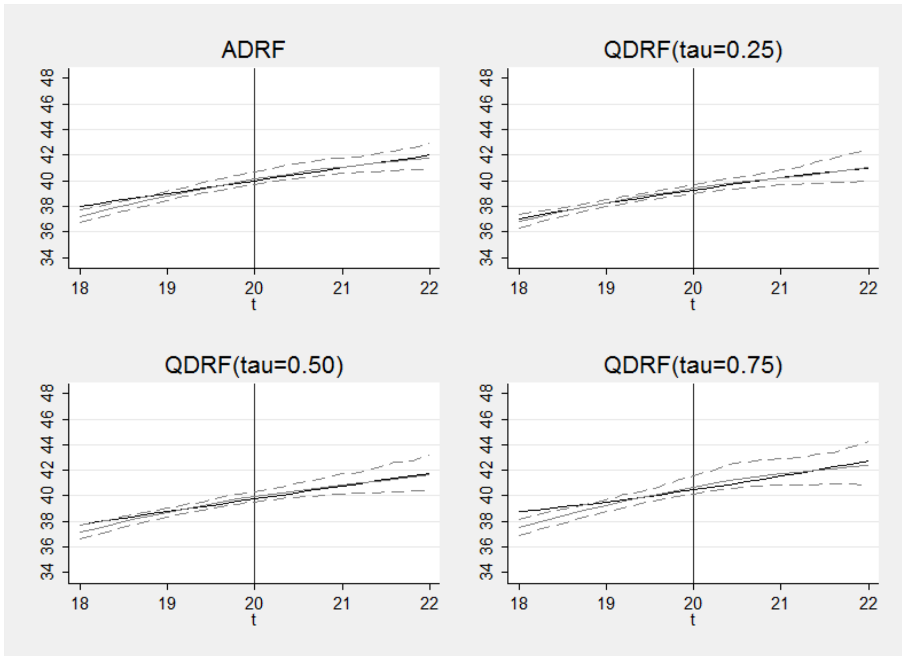
$n = 5000$



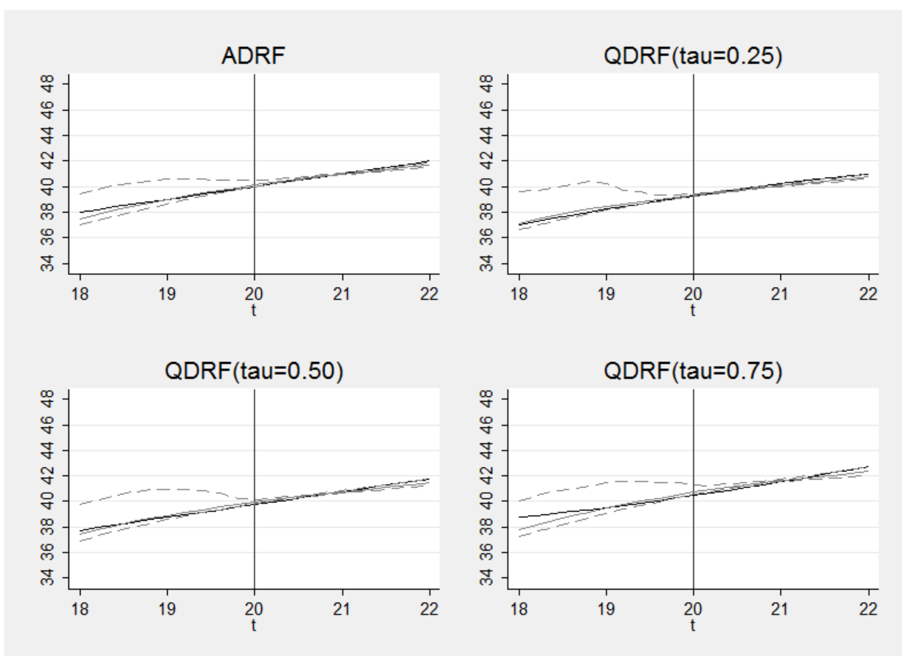
Notes: Monte Carlo experiments based on 500 simulations. In each figure the solid black line is the true value, the solid grey line correspond to the average estimated value of those functions at every t, and the dashed lines the 95% confidence intervals constructed from the Monte Carlo simulations.

Figure 7: DRF: Monte Carlo simulations, location-scale shift model, DGP: $\frac{\chi^2_3 - 3}{\sqrt{6}}$, Box-Cox model

$n = 100$



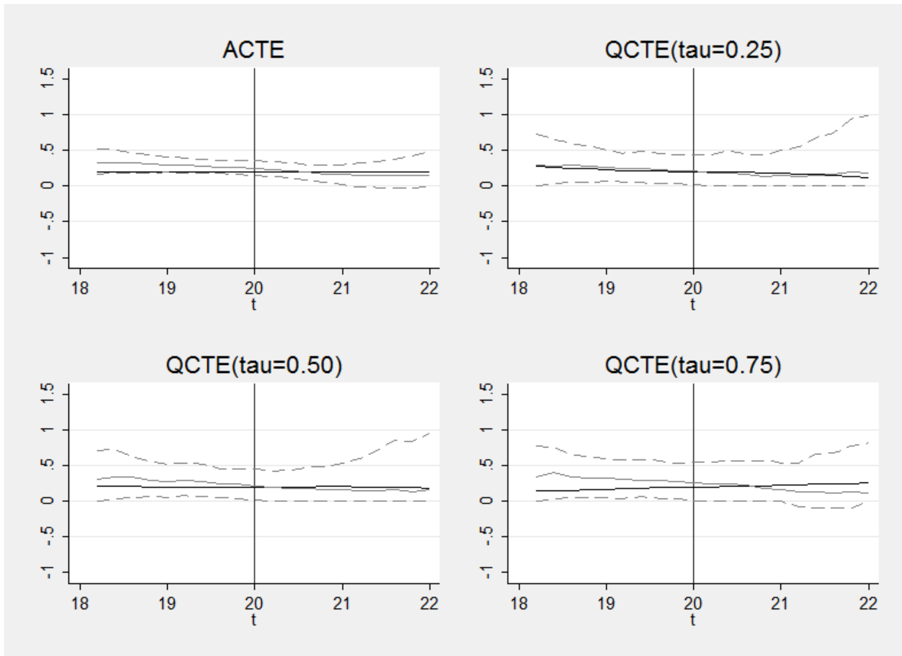
$n = 5000$



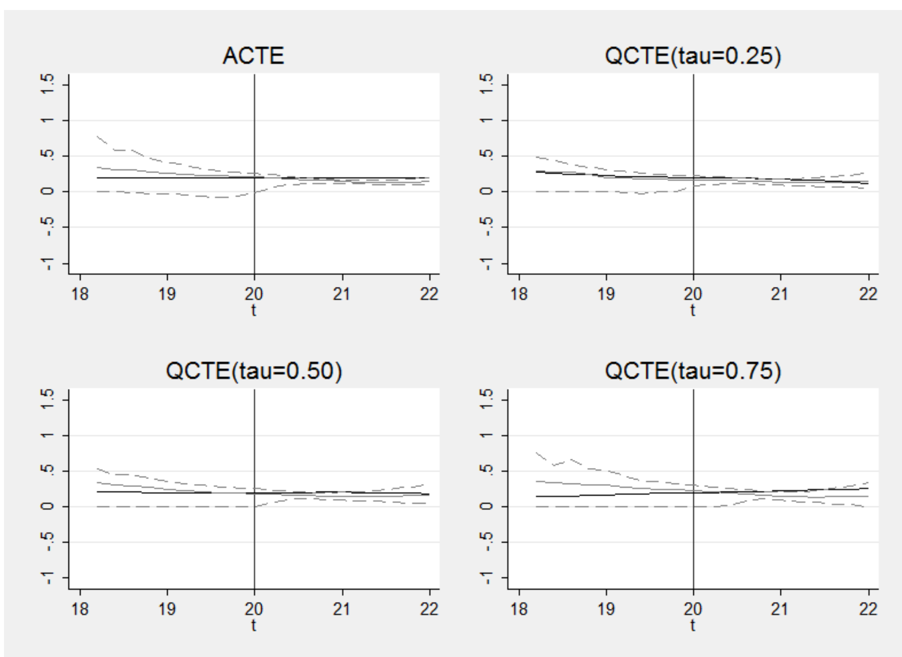
Notes: Monte Carlo experiments based on 500 simulations. In each figure the solid black line is the true value, the solid grey line correspond to the average estimated value of those functions at every t, and the dashed lines the 95% confidence intervals constructed from the Monte Carlo simulations.

Figure 8: CTE: Monte Carlo simulations, location-scale shift model, DGP: $\frac{\chi_3^2 - 3}{\sqrt{6}}$, Box-Cox model

$n = 100$



$n = 5000$



Notes: Monte Carlo experiments based on 500 simulations. In each figure the solid black line is the true value, the solid grey line correspond to the average estimated value of those functions at every t, and the dashed lines the 95% confidence intervals constructed from the Monte Carlo simulations.

Table 2: Monte Carlo experiments, location-scale shift model. DGP: $\frac{\chi_3^2-3}{\sqrt{6}}$.

n	Method	ADRF			QDRF25			QDRF50			QDRF75		
		MISE	ISBias	IVar	MISE	ISBias	IVar	MISE	ISBias	IVar	MISE	ISBias	IVar
100	Gaussian	1.002	0.057	0.945	1.175	0.051	1.124	1.573	0.148	1.425	1.924	0.248	1.676
100	Box-Cox	0.490	0.089	0.401	0.499	0.008	0.492	0.624	0.050	0.575	0.895	0.213	0.682
500	Gaussian	3.373	0.768	2.604	4.065	0.558	3.507	4.746	1.092	3.654	5.046	1.523	3.524
500	Box-Cox	0.246	0.099	0.148	0.171	0.009	0.162	0.287	0.054	0.233	0.514	0.221	0.293
1000	Gaussian	7.770	3.320	4.450	9.827	3.105	6.722	10.250	4.235	6.015	9.692	4.685	5.007
1000	Box-Cox	0.232	0.082	0.150	0.175	0.012	0.163	0.284	0.043	0.240	0.464	0.191	0.274
5000	Gaussian	16.058	10.837	5.221	20.489	12.830	7.659	19.351	12.983	6.369	16.823	11.257	5.565
5000	Box-Cox	0.245	0.039	0.205	0.298	0.026	0.272	0.320	0.025	0.295	0.415	0.119	0.295

Notes: Monte Carlo experiments based on 500 simulations.

because the lottery prize is a continuous treatment variable.

Although the lottery prize is obviously randomly assigned, there is substantial correlation between some of the background variables and the lottery prize in our sample. The main source of potential bias is the unit and item nonresponse. In the survey unit nonresponse was about 50%. In fact it was possible to directly demonstrate that this onresponse was nonrandom, since for all units the lottery prize was observed.Imbens et al. (2001) show that the higher the lottery prize, the lower the probability of responding to the survey. The missing data imply that the amount of the prize is potentially correlated with background characteristics and potential outcomes. In order to remove such biases we make the weak unconfoundedness assumption, that conditional on covariates the lottery prize is independent of the potential outcomes.

Imbens et al. (2001) use a simple framework to give an economic interpretation to their econometric specification. This setup assumes that each individual i maximizes a Stone-Geary utility function choosing leisure and consumption intertemporally throughout their life, subject to their intertemporal budget. The first order conditions of this problem results in a linear characterization of the optimal labor earnings as a function of the lottery prize:

$$Y_i = \beta\bar{\lambda}T_i + \theta' X_i + \varepsilon_i,$$

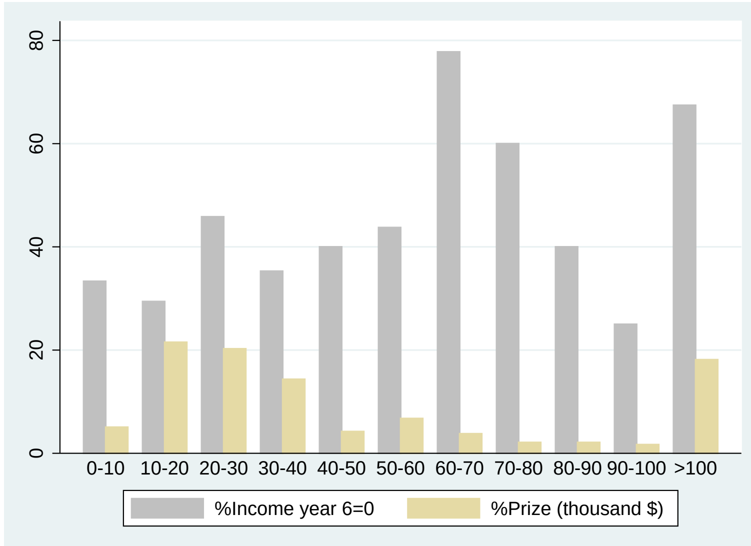
where X_i are additional covariates that correct for nonresponses bias and the residual ε_i captures others unobservable variations affecting non-labor income. The key assumption for ε_i is unconfoundedness. Note that the effect of the lottery prize on labor earnings is affected by the average factor $\bar{\lambda}$. This coefficient is a parameter that reflects the individual differences in lifetime, as well as variability between the discount factor and the interest rate. Therefore, this simple framework gives us an interesting interpretation of the CTE heterogeneity for each quantile beyond the mean. In particular, quantile analysis could be interpreted in terms of λ heterogeneity.

The sample we use in this analysis is the "winners" sample of 237 individuals who won a major prize in the lottery. For each individual we observe social security earnings for six years before the lottery and six years after. The outcome of interest is `year6` (earnings six years after winning the lottery), denoted Y, and the treatment is `prize`, the prize amount, denoted T. Control variables X are age, gender, years of high school, years of college, winning year, number of tickets bought, work status after winning, and earnings s years before winning the lottery (with $s = 1, 2, ..., 6$). Of these 237 individuals we keep a sample of 202 for whom we have income information on income Y. Detailed descriptive statistics can be found in Imbens et al. (2001) and Hirano and Imbens (2004).

As in the Monte Carlo section we estimate the model with two different approaches. First, following Hirano and Imbens (2004), we estimate a log-linearized models of the prize amount (being a non-zero variable) as a function of either X or (X, Y), assuming that the conditional distribution of $log(T)$ is normal. Second, we implement the flexible Box-Cox model with different parameters for $log(T)$ and the explanatory variables.

We consider $\mathcal{T} = (0, 100]$ (in thousands of US dollars) and a grid $t \in \{10, 20, ..., 100\}$. Moreover we consider marginal increments of 10 (i.e. \$10,000 dollars). A feature of the data to be considered is that almost half the sample has $Y = 0$ (52% which corresponds to

Figure 9: Distribution of lottery prices and no income 6 years later



47% for male and 59% for female). That is, half the sample is not working and receive no income 6 years after winning the lottery. Figure 9 reports the distribution of prize values and the percentage of zeros for the dependent variable. We follow Hirano and Imbens (2004); Bia and Mattei (2008) approach who considers that a zero value correspond to an observed level of income and it requires no truncation analysis. In fact, a value of $Y = 0$ informs us that the individual is not working, which is an important issue for the analysis of the effect of non-labor income on the labor force. We find that for low quantiles, i.e. $\tau < 0.5$, $QDRF_\tau(t) = 0, \forall t \in \mathcal{T}$. We then only report the QRDF for $\tau \in [0.50, 1)$. Note that the existence of a non-negligible mass at $Y = 0$ does not invalidate asymptotic inference on the QDRF estimator, provided that the density of the outcome at 0 is positive.

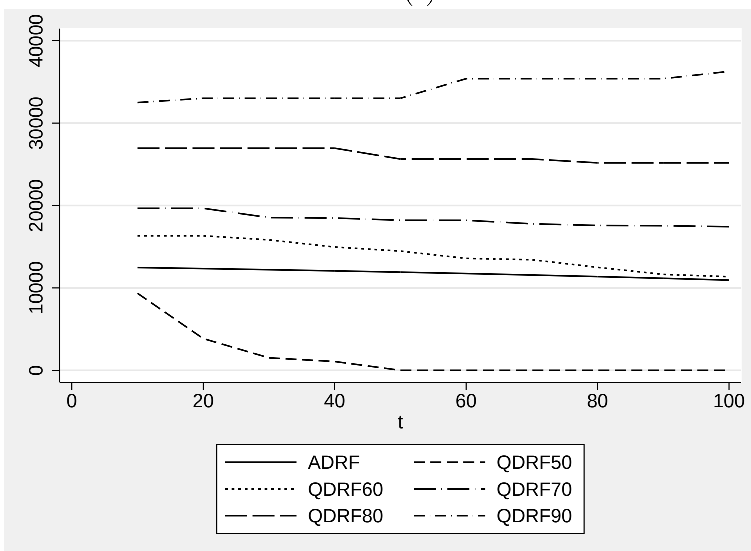
We first report ADRF and QRDF functions when we impose no correction for participation, i.e. $w(\cdot) = 1$. In particular we estimate a locally weighted kernel regression procedure, with weights given by a normal density of $log(T)$ evaluated at a local mean of $log(t)$ and standard deviation of 1 ($log(T)$ has a sample mean of 3.59 and a standard deviation of 0.95). Figure 10 (left panel) reports the QDRF and figure 10 (right panel) shows the CTE for increments of 10, for both average and quantiles, i.e. $CTE(t, t-10) = DRF(t)-DRF(t-10), \forall t \in \mathcal{T}$. DRF analysis suggests that the lottery prize has no effect on labor income. That is, receiving a larger prize has no effect on labor earnings, and this applies uniformly for all quantiles. Moreover, the quantiles only change the location but not the scale, that is, they are parallel lines.

Next consider our proposed two-step estimator. Figure 11 (top figure) reports the ADRF as in Galvao and Wang (2015) together with the QDRF for selected quantiles, for the Box-Cox model only. The graph shows an ordered cascade, where $Y(t)$ looks as a decreasing function of t, contrary to the uncorrected QRDF functions in figure 10. The results show that the non-responding pattern described above masks the fact that non-labor income has a negative effect on labor income for the entire distribution, upper quantiles included. Note that the figure suggests a particular pattern in which there is a prize threshold value for which the income becomes zero. That is, for all $\tau \in (0,1)$ there exist a value t_τ such that $QDRF_\tau(t) = 0, \forall t > t_\tau \in \mathcal{T}$ and $t_{\tau_1} \geq t_{\tau_2}$ for $\tau_1 > \tau_2,\ \tau_1, \tau_2 \in (0,1)$. This suggests that there is always a prize threshold value that is high enough to make all individuals stop working, and that this applies uniformly for all quantiles. This nonlinearity in the ADRF is explored in Imbens et al. (2001) using a quadratic specification and non-parametrically in Hirano and Imbens (2004). Both estimates shows a convex relationship suggesting a marginally decreasing effect of the lottery price on labor earnings. Our contribution to these estimates is that this convex behavior is homogeneous in the rest of the labor earnings distribution and then the threshold value is monotonic in the quantiles. Table 3 presents a formal test for the null hypothesis of income being zero 6 years after the prize uniformly over the prize values. The results show that we reject the null hypothesis for mean and $\tau \geq 0.6$, but cannot reject for $\tau = 0.5$.

Figure 11 (bottom figure) shows the estimates of the continuous TE of increments of 10, for both average and quantiles, i.e. $CTE(t, t-10) = DRF(t) - DRF(t-10), \forall t \in \mathcal{T}$. The figure shows that for low values of the prize ($t < 40$) an increase in the prize reduces income 6 years later. In addition, this effect seems deeper for the 50-70 quantiles but less sharp for the upper quantiles. However as the prize move to higher values ($t \geq 40$), the

Figure 10: Uncorrected average and quantile ($\tau \in (0.5, 0.6, 0.7, 0.8, 0.9)$) DRF and CTE. Changes of U$S 10k

$DRF(t)$



$CTE(t, t-10) = DRF(t) - DRF(t-10)$

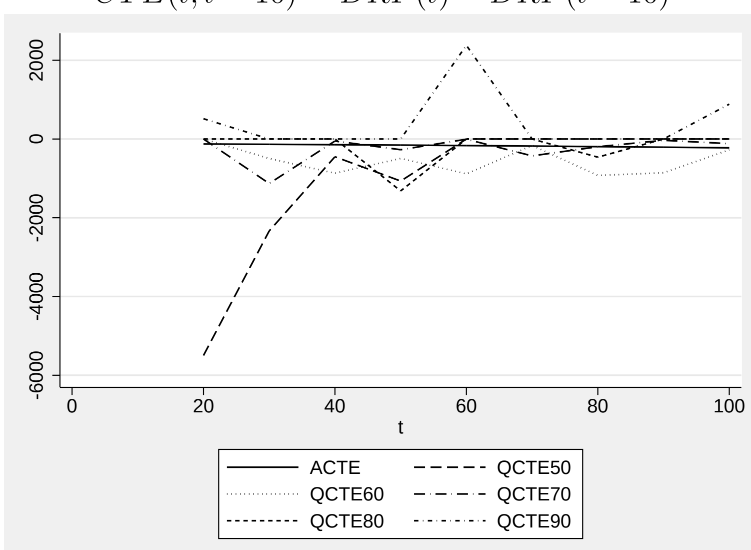


Table 3: Hypothesis: $DRF(t) = 0$

	T_{1n}	$C_{1,0.95}$	p-value	T_{2n}	$C_{2,0.95}$	p-value
ADRF	240840.5	82363.7	0.014	152175.5	49767.5	0.020
QDRF50	236456.2	236456.2	0.022	39796.9	45388.2	0.064
QDRF60	307917.5	213289.5	0.018	114137.7	114137.7	0.036
QDRF70	397201.5	212650.0	0.004	243933.5	118518.8	0.014
QDRF80	461854.9	141138.9	0.000	328012.8	85391.1	0.002
QDRF90	590508.0	141373.4	0.000	464208.6	91127.4	0.002

Note: $T_{1n} = \sqrt{n} \sup_{t \in \mathcal{T}} |\widehat{q}_\tau(t) - r(t)|$ and $T_{2n} = \sqrt{n} \int_{t \in \mathcal{T}} |\widehat{q}_\tau(t) - r(t)| \, dt$. Critical values and p-values computed by bootstrap (500 replications).

marginal effect of the prize vanishes, i.e. it does not affect the income 6 years later. Table 4 presents a formal test for the null hypothesis of the effect of increasing the prize value by 10 has no effect on income 6 years after the prize uniformly over the prize values. We compute the test statistics for all t and for prize subsamples on each side of the observed prize threshold ($t = 40$). The results show that we reject the null hypothesis for the mean in all the cases, but we find mixed evidence for quantile effects. Note that although sampling errors do not allow rejection of the null hypothesis in all cases, p-values appear larger when $t > 40$ indicating that the marginal effect of the prize on earnings is weaker.

Figure 11: Average and quantile ($\tau \in (0.5, 0.6, 0.7, 0.8, 0.9)$) DRF and CTE. Changes of U$S 10k. Box-Cox model.

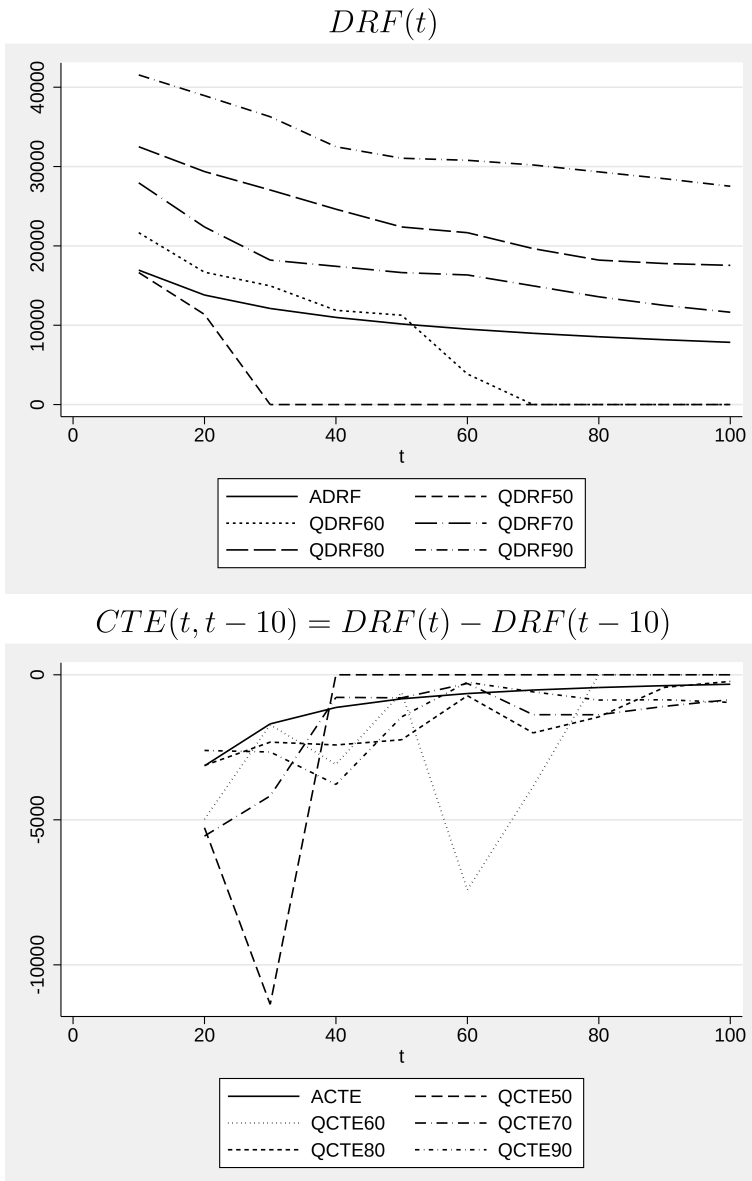


Table 4: Hypothesis: $CTE(t) = 0$

	T_{3n}	$C_{3,0.95}$	p-value	T_{4n}	$C_{4,0.95}$	p-value
			All prizes			
ACTE	44605.8	33298.7	0.018	14371.0	8689.3	0.016
QCTE50	161512.8	161512.8	0.038	26272.9	36365.5	0.459
QCTE60	105387.0	141494.2	0.327	34213.1	44101.9	0.427
QCTE70	79022.4	111185.7	0.156	25753.4	30391.4	0.106
QCTE80	44443.0	86633.3	0.343	23604.1	23787.3	0.050
QCTE90	53752.3	78397.1	0.240	22157.6	24065.2	0.124
			Low prizes ($t \leq 40$)			
ACTE	44605.8	33298.7	0.018	28228.4	18135.1	0.014
QCTE50	161512.8	161512.8	0.034	78818.7	106258.7	0.403
QCTE60	70779.1	112429.3	0.150	46375.9	58127.5	0.120
QCTE70	79022.4	79022.4	0.016	49820.1	47304.5	0.016
QCTE80	44443.0	86512.5	0.315	37279.8	42775.4	0.102
QCTE90	53752.3	73849.0	0.170	42884.4	45859.6	0.080
			High prizes ($t > 40$)			
ACTE	11780.6	5980.0	0.014	7442.3	3717.8	0.012
QCTE50	0.0	93888.9	0.164	0.0	22148.1	0.164
QCTE60	105387.0	138573.5	0.275	28131.6	51466.4	0.577
QCTE70	19585.1	108655.9	0.385	13720.0	31263.1	0.275
QCTE80	31836.4	43675.5	0.204	16766.2	18779.7	0.114
QCTE90	20594.2	59039.4	0.537	11794.1	19386.1	0.409

Note: $T_{3n} = \sqrt{n}\sup_{t\in\mathcal{T}} |\widehat{\Delta}_\tau(t, t+\delta) - r(t)|$ and $T_{4n} = \sqrt{n}\int_{t\in\mathcal{T}} |\widehat{\Delta}_\tau(t, t+\delta) - r(t)|\,dt$. Critical values and p-values computed by bootstrap (500 replications).

The application illustrates that the new method is an important tool to study continuous TE. Our empirical results document strong heterogeneity on the responses of non-labor income changes on subsequent labor earnings. In fact our results complement the findings of Hirano and Imbens (2004) where larger prizes were associated with lower labor earnings. The quantile analysis also reveals that larger prizes produce lower labor earnings, but a larger prize is required for individuals in the upper part of the distribution of unobservables.

8. Conclusion

This paper studies quantile treatment effects models with a continuous treatment by developing identification, and practical estimation and inference. The identification of the parameters of interest, the dose-response functions and the quantile treatment effects, is based on the assumption that selection to treatment is based on observable characteristics. This assumption is largely employed in the literature for identification of treatment effects.

We suggest an easy way to implement in practice semiparametric two-step estimator for the parameters of interest, and establish its asymptotic properties. Importantly, we develop practical statistical inference procedures and establish the theoretical validity of a bootstrap approach to implement these methods in applications and for a wide range of testing procedures.

Monte Carlo simulations show that the proposed methods have good finite sample properties. The experiments show that, for a relatively small sample size, the method produces estimates with good precision and negligible bias, especially for middle quantiles. Finally, we apply the proposed methods to estimate the unconditional quantile effects of the effect of the prize amount on subsequent labor earnings. Results from the empirical application indicate that the method works relatively well in practice and reveals that the effect of non-labor income has a monotonic effect on labor earnings. Moreover, there is considerable heterogeneity in the upper quantiles of the distribution of earnings.

An important extension of the methods developed in this paper is to apply it for weak convergence on income inequalities and other functionals that one can construct from quantile processes (e.g., via Hadamard differentiable functionals). For instance this could be generalized to Firpo and Pinto (2016) analysis of inequality measures in the continuous context or to maximum treatment effects.

Acknowledgements

The authors would like to express their appreciation to the Co-Editor Erricos John Kontoghiorghes, the Associate Editor, two anonymous referees, Sergio Firpo, Alfonso Flores-Lagunes, Roger Koenker, Andrea Rotnitzky, and Yuya Sasaki for helpful comments and discussions. All the remaining errors are ours. This research has been partially supported by grants ECO2013-46516-C4-1-R (Equity and Poverty: Methods and Implications, Ministerio de Economía y Competitividad, Gobierno de España) and SGR2014-1279 (Equity and Development Research Group, Generalitat de Catalunya).

Abadie, A., Angrist, J., Imbens, G., 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. Econometrica 70, 91–117.

Abadie, A., Imbens, G., 2006. Large sample properties of matching estimators for average treatment effects. Econometrica 74, 235–267.

Angrist, J., Chernozhukov, V., Fernandez-Val, I., 2006. Quantile regression under misspecification, with an application to the u.s. wage structure. Econometrica 74, 539–563.

Bia, M., Flores, C. A., Flores-Lagunes, A., Mattei, A., 2014. A stata package for the application of semiparametric estimators of dose-response functions. Stata Journal 14(3), 580–604.

Bia, M., Mattei, A., 2008. A stata package for the estimation of the dose-response function through adjustment for the generalized propensity score. Stata Journal 8, 354–373.

Carroll, R. J., Ruppert, D., 1984. Power transformation when fitting theoretical models to data. Journal of the American Statistical Association 79, 321–328.

Cattaneo, M., 2010. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. Journal of Econometrics 155, 138–154.

Cattaneo, M., Drukker, D. M., Holland, A. D., 2013. Estimation of multivalued treatment effects under conditional independence. Stata Journal 13, 407–450.

Chen, X., Linton, O., Van Keilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. Econometrica 71, 1591–1608.

Chernozhukov, V., Hansen, C., 2005. An iv model of quantile treatment effects. Econometrica 73, 245–261.

Chernozhukov, V., Hong, H., 2002. Three-step censored quantile regression and extramarital affairs. Journal of the American Statistical Association 97, 872–882.

Cohen, J., 1983. The cost of dichotomization. Applied Psychological Measurement 7, 249–253.

Cox, D. R., 1957. Note on grouping. Journal of the American Statistical Association 52, 543–547.

Dehejia, R., Wahba, S., 1999. Causal effects in nonexperimental studies: Revaluating the evaluation of training programs. Journal of the American Statistical Association 94, 1053–1062.

d'Haultfoeuille, X., Hoderlein, S., Sasaki, Y., 2015. Nonlinear difference-in-difference in repeated cross sections with continuous treatments, mimeo.

Fedorov, V., Mannino, F., Zhang, R., 2009. Consequences of dichotomization. Pharmaceutical Statistics 8, 50–61.

Firpo, S., 2007. Efficient semiparametric estimation of quantile treatment effects. Econometrica 75, 259–276.

Firpo, S., Pinto, C., 2016. Identification and estimation of distributional impacts of interventions using changes in inequality measures. Journal of Applied Econometrics 31, 457486.

Florens, J. P., Heckman, J. J., Meghir, C., Vytlacil, E., 2008. Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. Econometrica 76, 1191–1206.

Flores, C. A., 2007. Estimation of dose-response functions and optimal doses with a continuous treatment, mimeo.

Flores, C. A., Flores-Lagunes, A., Gonzalez, A., Neumann, T., 2012. Estimating the effects of length of exposure to instruction in a training program: The case of job corps. Review of Economics and Statistics 94, 153–171.

Galvao, A., Wang, L., 2015. Uniformly semiparametric efficient estimation of treatment effects with a continuous treatment. Journal of the American Statistical Association 110, 1528–1542.

Hahn, J., 1998. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. Econometrica 66, 315–331.

Heckman, J., Ichimura, H., Smith, J., Todd, P., 1998. Characterizing selection bias using experimental data. Econometrica 66, 1017–1098.

Hirano, K., Imbens, G., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71, 1161–1189.

Hirano, K., Imbens, G. W., 2004. The propensity score with continuous treatment. In: Gelman, A., Meng, X.-L. (Eds.), Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives. Wiley.

Imai, G., van Dyk, D., 2004. Causal inference with general treatment regimes: Generalizing the propensity score. Journal of the American Statistical Association 99, 854–866.

Imbens, G., 2000. The role of the propensity score in estimating dose-response function. Biometrika 87, 706–710.

Imbens, G., Newey, W., Ridder, W., 2006. Mean-squared-error calculations for average treatment effects, mimeo.

Imbens, G., Rubin, D., Sacerdote, B., 2001. Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. American Economic Review 91(4), 778–794.

Kennedy, E. H., Ma, Z., McHugh, M. D., Small, D. S., 2015. Nonparametric methods for doubly robust estimation of continuous treatment effects. Https://arxiv.org/abs/1507.00747.

Koenker, R., Bassett, G. W., 1978. Regression quantiles. Econometrica 46, 33–49.

Kosorok, M., 2008. Introduction to Empirical Processes and Semiparametric Inference. Springer.

Lechner, M., 2001. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In: Lechner, M., Pfeiffer, F. (Eds.), Econometric Evaluation of Labour Market Policies. Physical Springer, Heidelberg.

Lee, Y., 2015. Partial mean processes with generated regressors: Continuous treatment effects and nonseparable models, mimeo.

Li, Q., Racine, J., Wooldridge, J., 2009. Efficient estimation of average treatment effects with mixed categorical and continuous data. Journal of Business & Economic Statistics 27, 206–223.

Montes-Rojas, G., 2009. A note on the variance of average treatment effects estimators. Economics Bulletin 29(4), 2937–2943.

Murphy, K., Topel, R., 1985. Estimation and inference in two-step econometric models. Journal of Business and Economic Statistics 3, 88–97.

Newey, W. K., McFadden, D. L., 1994. Large sample estimation and hypothesis testing. In: Engle, R. F., McFadden, D. L. (Eds.), Handbook of Econometrics, Vol. 4. North Holland, Elsevier, Amsterdam.

Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Rothe, C., 2010. Nonparametric estimation of distributional policy effects. Journal of Econometrics 155(1), 56–70.

Rubin, D., 1977. Assignment to treatment group on the basis of a covariate. Journal of Educational Statistics 2, 1–28.

van Belle, G., 2008. Statistical Rules of Thumb. Wiley-Interscience; 2nd edition.

van der Vaart, A., 1998. Asymptotic Statistics. Cambridge University Press, New York, New York.

Wand, M., Marron, J., Ruppert, D., 1991. Transformations in density estimation. Journal of the American Statistical Association 86, 343–353.

Wang, N., Ruppert, D., 1995. Nonparametric estimation of the transformation in the transform-both-sides regression model. Journal of the American Statistical Association 90, 522–534.

Wei, Y., Carroll, R. J., 2009. Quantile regression with measurement error. Journal of the American Statistical Association 104, 1129–1143.

9. Appendix

Notations: Let E and $\mathbb{E}_n$ 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 $|g(x)|_\infty$ denote $\sup_x |g(x)|$ for $x \in \mathcal{X} \subset \mathbb{R}^d$, where d is a positive integer.

Proof of Lemma 1. Fixing $t = t_0$, by law of iterated expectations, $\mathrm{E}[m(Y(t_0); q_\tau(t_0))] = \mathrm{E}\{\mathrm{E}[m(Y(t_0); q_\tau(t_0))|\boldsymbol{X}]\}$. For the conditional expectation,

$$\begin{aligned}\mathrm{E}[m(Y(t_0); q_\tau(t_0))|\boldsymbol{X}] =& \mathrm{E}[m(Y(t_0); q_\tau(t_0))|\boldsymbol{X}, T = t_0] = \mathrm{E}[m(Y; q_\tau(t_0))|\boldsymbol{X}, T = t_0] \\ =& \lim_{\delta t \downarrow 0} \mathrm{E}[m(Y; q_\tau(t_0))|\boldsymbol{X}, T \in [t_0, t_0 + \delta t]],\end{aligned}$$

where the first equality is by condition **I.1**, the second equation is by the fact that if $T = t_0$, then $Y = Y(t_0)$, and the third equality is by condition **I.2**. Moreover, we have

$$\mathrm{E}[m(Y; q_\tau(t_0))|\boldsymbol{X}, T \in [t_0, t_0 + \delta t]] = \mathrm{E}[\mathbf{1}(T \in [t_0, t_0 + \delta t])m(Y; q_\tau(t_0))|\boldsymbol{X}, T \in [t_0, t_0 + \delta t]]. \quad (9)$$

By law of total expectation

$$\begin{aligned}&\mathrm{E}[\mathbf{1}\{T \in [t_0, t_0 + \delta t]\}m(Y; q_\tau(t_0))|\boldsymbol{X}] \\ =&\mathrm{E}[\mathbf{1}\{T \in [t_0, t_0 + \delta t]\}m(Y; q_\tau(t_0))|\boldsymbol{X}, T \in [t_0, t_0 + \delta t]]\mathrm{P}(T \in [t_0, t_0 + \delta t]|\boldsymbol{X}) \\ +&\mathrm{E}[\mathbf{1}\{T \in [t_0, t_0 + \delta t]\}m(Y; q_\tau(t_0))|\boldsymbol{X}, T \notin [t_0, t_0 + \delta t]]\mathrm{P}(T \notin [t_0, t_0 + \delta t]|\boldsymbol{X}),\end{aligned}$$

the right hand side of equation (9) equals

$$\frac{\mathrm{E}[\mathbf{1}\{T \in [t_0, t_0 + \delta t]\}m(Y; q_\tau(t_0))|\boldsymbol{X}]}{\mathrm{P}(T \in [t_0, t_0 + \delta t]|\boldsymbol{X})} = \frac{\mathrm{E}[(\mathbf{1}\{T \leq t_0 + \delta t\} - \mathbf{1}\{T < t_0\})m(Y; q_\tau(t_0))|\boldsymbol{X}]}{F_{T|\boldsymbol{X}}(t_0 + \delta t|\boldsymbol{X}) - F_{T|\boldsymbol{X}}(t_0|\boldsymbol{X})},$$

where $F_{T|\boldsymbol{X}}$ denotes the conditional distribution function of T given $\boldsymbol{X}$. Noting that

$$\begin{aligned}&\mathrm{E}[(\mathbf{1}\{T \leq t_0 + \delta t\} - \mathbf{1}\{T \leq t_0\})m(Y; q_\tau(t_0))|\boldsymbol{X} = \boldsymbol{x}] \\ =&\iint (\mathbf{1}\{t \leq t_0 + \delta t\} - \mathbf{1}\{t \leq t_0\})m(y; q_\tau(t_0))f_{T,Y|\boldsymbol{X}}(t, y|\boldsymbol{x})\, dtdy \\ =&\int \int_{t_0}^{t_0+\delta t} m(y; q_\tau(t_0))f_{T,Y|\boldsymbol{X}}(t, y|\boldsymbol{x})\, dtdy,\end{aligned}$$

it follows that

$$\begin{aligned}\lim_{\delta t \downarrow 0} \mathrm{E}[m(Y; q_\tau(t_0))|\boldsymbol{X} = \boldsymbol{x}, T \in [t_0, t_0 + \delta t]] =& \lim_{\delta t \downarrow 0} \frac{\int \int_{t_0}^{t_0+\delta t} m(y; q_\tau(t_0))f_{T,Y|\boldsymbol{X}}(t, y|\boldsymbol{x})\, dtdy}{F_{T|\boldsymbol{X}}(t_0 + \delta t|\boldsymbol{x}) - F_{T|\boldsymbol{X}}(t_0|\boldsymbol{x})} \\ =& \lim_{\delta t \downarrow 0} \frac{\int m(y; q_\tau(t_0))f_{T,Y|\boldsymbol{X}}(t_0 + \epsilon_1 \delta t, y|\boldsymbol{x})\, dy}{f_{T|\boldsymbol{x}}(t_0 + \epsilon_2 \delta t|\boldsymbol{x})} \\ =& \frac{\int m(y; q_\tau(t_0))f_{T,Y|\boldsymbol{X}}(t_0, y|\boldsymbol{x})\frac{f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})}{f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})}\, dy}{f_{T|\boldsymbol{X}}(t_0|\boldsymbol{x})} \\ =& \frac{\int m(y; q_\tau(t_0))f_{T|\boldsymbol{X},Y}(t_0|\boldsymbol{x}, y)f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\, dy}{f_{T|\boldsymbol{X}}(t_0|\boldsymbol{x})} \\ =& \frac{\mathrm{E}[m(Y; q_\tau(t_0))f_{T|\boldsymbol{X},Y}(t_0|\boldsymbol{x}, Y)|\boldsymbol{X} = \boldsymbol{x}]}{f_{T|\boldsymbol{X}}(t_0|\boldsymbol{x})},\end{aligned}$$

where ϵ_1 and ϵ_2 are fixed numbers in [0,1]. The second equality is by mean value theorems for differentiation and integration. And the third equality is by the assumptions on the theorem and dominated convergence theorem. Hence $\mathrm{E}[m(Y(t_0); q_\tau(t_0))] = \mathrm{E}\left[m(Y; q_\tau(t_0)) \frac{f_{T|Y,\boldsymbol{X}}(t_0|Y,\boldsymbol{X})}{f_{T|\boldsymbol{X}}(t_0|\boldsymbol{X})}\right]$. ■

Proof of Corollary 1. The result follows directly from Lemma 1 by observing that $q_\tau(t)$ is identified by each given t. Thus, $\Delta_\tau(t,t')$ is also identified. ■

Proof of Proposition 1.

Assume conditions **A.1** and **A.2**. Recall that $\mu := (\lambda_1, \lambda_2, \boldsymbol{\beta}_X, \beta_Y, \sigma_\epsilon^2)$. Under **A.1** and **A.2**, $w = w(\boldsymbol{u}; t; \mu)$, where $\mu \in \mathbb{R}^{d_\mu}$ with d_μ being a positive integer, and $w(\boldsymbol{u}; t; \mu)$ is a smooth function of μ with uniformly continuous, bounded, and square integrable first derivative, $w'(\boldsymbol{u}; t; \mu)$, with respect to μ. Now, by the standard properties of the MLE estimator, $\sqrt{n}(\widehat{\mu} - \mu) \xrightarrow{d} N(0, \Im_\mu^{-1})$ where $\Im_\mu$ is nonsingular.

Given **A.1** and **A.2**, for conditions **PC.1** and **PG.1**, by Theorem 19.7 of van der Vaart (1998), this parametric Lipschitz continuous functional class is Donsker. From **A.1** and **A.2**, implication **PC.2** is direct.

By Mean Value Theorem, $w(\boldsymbol{u}; \cdot; \widehat{\mu}) - w(\boldsymbol{u}; \cdot; \mu_0) = w'(\boldsymbol{u}; \cdot; \mu^*)(\widehat{\mu} - \mu_0)$, where μ^* is a convex combination of $\widehat{\mu}$ and μ_0. Therefore,

$$|w(\boldsymbol{u}; \cdot; \widehat{\mu}) - w(\boldsymbol{u}; \cdot; \mu_0)|_\infty = |w'(\boldsymbol{u}; \cdot; \mu^*)(\widehat{\mu} - \mu_0)|_\infty \leq |w'(\boldsymbol{u}; \cdot; \mu^*)|_\infty ||\widehat{\mu} - \mu_0|| = O_p(n^{-1/2})$$

since $|w'(\boldsymbol{u}; \cdot; \mu^*)|_\infty$ is bounded and $||\widehat{\mu} - \mu_0|| = O_p(n^{-1/2})$. There, conditions **PC.3** and **PG.3** are verified.

To verify the weak convergence of condition **PG.2**, we need to use the functional delta method, which involves Hadamard differentiability of a map between norm spaces. A map $\phi : \mathbb{D}_\phi \mapsto \mathbb{E}_n$ is Hadamard differentiable at $\theta \in \mathbb{D}$, tangentially to a set $\mathbb{D}_0$, if there exists a continuous linear map $\phi'_\theta : \mathbb{D} \mapsto \mathbb{E}_n$ such that $\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \to \phi'_\theta(h)$, as $n \to \infty$, for all converging sequences $t_n \to 0$ and $h_n \to h \in \mathbb{D}_0$, with $h_n \in \mathbb{D}$ and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all $n \geq 1$ sufficiently large; see p. 22 of Kosorok (2008). We first verify that w is Hadamard differentiable at μ tangentially to $\mathbb{R}^{d_\mu}$. For any $l_n \to 0$ and $h_n \to h \in \mathbb{R}^{d_\mu}$,

$$\frac{w(\boldsymbol{u}; t; \mu + l_n h_n) - w(\boldsymbol{u}; t; \mu)}{l_n} = \frac{w'(\boldsymbol{u}; t; \mu^*) l_n h_n}{l_n} \to w'(\boldsymbol{u}; t; \mu) h.$$

Using the functional delta method, since $w'(\boldsymbol{u}; t; \mu)$ is uniformly bounded,

$$\sqrt{n}(w(\boldsymbol{u}; t; \widehat{\mu}) - w(\boldsymbol{u}; t; \mu)) \rightsquigarrow w'(\boldsymbol{u}; t; \mu) Z_\mu \text{ in } \ell^\infty(\mathcal{U} \times \mathcal{T})$$

where $Z_\mu \sim N(0, \Im_\mu^{-1})$.

■

Proof of Theorem 1. The general result for consistency of Z-estimators is given in **Lemma 2** in the Online Supplemental Appendix. To prove the result we apply the lemma to our continuous treatment model with $\theta_0 = q_0(\cdot)$, $h_0 = w_0(\cdot)$, $\mathbb{Z}(\theta, h)(t) = \mathbb{E}_n\psi_{q,w,t}$, and $\mathrm{Z}(\theta, h)(t) = \mathrm{E}\psi_{q,w,t}$, where $\psi_{q,w,t} = m(y; q_\tau(t))w(\boldsymbol{U}; t) = (\tau - \mathbf{1}\{Y < q_\tau(t)\})w(\boldsymbol{U}; t)$.

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 $\mathcal{U}\times\mathcal{T}$, and $||\cdot||_\mathcal{H} = ||\cdot||_\Pi = \sup_{\boldsymbol{u}\in\mathcal{U}} |\cdot| = |\cdot|_\infty$. For any $\delta > 0$, $\Pi = \{w \in \Pi : |w - w_0|_\infty < \delta\}$.

To establish the result we verify the conditions of **Lemma 2**. Thus, under **PC.1**–**PC.4** 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 **PC.2** and **PC.3** such that we have

$$
\begin{aligned}
|\mathbb{E}_n[(\tau - \mathbf{1}\{Y < \widehat{q}_\tau(\cdot)\})\widehat{w}(\boldsymbol{U}; \cdot)]| &\leq const \cdot \sup_{i\leq n} \frac{\widehat{w}(\boldsymbol{U}_i; \cdot)}{n} \\
\leq & const \cdot \frac{||\widehat{w}(\boldsymbol{U}; \cdot)||_\Pi}{n} = const \cdot \frac{||w_0(\boldsymbol{u}; \cdot)||_\Pi + o_p(1)}{n} = O_{p^*}(1/n).
\end{aligned}
$$

Condition **C.2** holds by condition **PC.4**.

We now show that condition **C.3**, the continuity of $\mathrm{E}[m(Y; q_\tau(t))w(\boldsymbol{U}; t)]$ at w_0 uniformly over $q_\tau(t) \in \ell^\infty(\mathcal{T})$, is satisfied. For any $||w - w_0||_\infty \leq \delta$, which is equivalent to $\sup_{t\in\mathcal{T}} \sup_{\boldsymbol{u}\in\mathcal{U}} |w(\boldsymbol{u}; t) - w_0(\boldsymbol{u}; t)| \leq \delta$, we have

$$
\begin{aligned}
&\sup_{t\in\mathcal{T}} \sup_{\boldsymbol{u}\in\mathcal{U}} |\mathrm{E}[m(Y; q_\tau(t))w(\boldsymbol{U}; t)] - \mathrm{E}[m(Y; q_\tau(t))w_0(\boldsymbol{U}; t)]| \\
=&\sup_{t\in\mathcal{T}} \sup_{\boldsymbol{u}\in\mathcal{U}} |\mathrm{E}[m(Y; q_\tau(t))(w(\boldsymbol{U}; t) - w_0(\boldsymbol{U}; t))]| \leq \sup_{t\in\mathcal{T}} |\mathrm{E}[m(Y; q_\tau(t))]|\delta.
\end{aligned}
$$

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

Finally, let $\psi_{q,1,t} = (\tau - \mathbf{1}\{Y < q_\tau(t)\})$ and note that the function class $\{\psi_{q,1,t} : q \in \ell^\infty(\mathcal{T}), t \in \mathcal{T}\}$ is Glivenko-Cantelli because it is a Vapnik-Červonenkis class. This result, condition **PC.1**, and Corollary 9.27 of Kosorok (2008) imply condition **C5S** which in turn implies **C.5**.

Hence, all the conditions of **Lemma 2** are satisfied. ■

Proof of Corollary 2. The result follows from Theorem 1, which establishes convergence in probability for each given t and the continuous mapping theorem. ■

Proof of Theorem 2. To establish the result we apply **Lemma 3** in the Online Supplemental Appendix and we verify its conditions.

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

$$
\begin{aligned}
&|\mathrm{E}[(\tau - \mathbf{1}\{Y \leq q_\tau(\cdot)\})w_0(\boldsymbol{U}; \cdot)] - \mathrm{E}[(\tau - \mathbf{1}\{Y \leq q_{\tau 0}(\cdot)\})w_0(\boldsymbol{U}; \cdot)] + \mathrm{E}[w_0(\boldsymbol{U}; \cdot)f_Y(q_{\tau 0})(q_\tau(\cdot) - q_{\tau 0}(\cdot))]|_\infty \\
=&|\mathrm{E}[\{\mathbf{1}\{Y \leq q_{\tau 0}(\cdot)\} - \mathbf{1}\{Y \leq q_\tau(\cdot)\} + f_Y(q_{\tau 0})(q(\cdot) - q_{\tau 0}(\cdot))\}w_0(\boldsymbol{U}; \cdot)]|_\infty \\
\asymp&|\mathrm{E}[\{\mathbf{1}\{Y \leq q_{\tau 0}(\cdot)\} - \mathbf{1}\{Y \leq q_\tau(\cdot)\} + f_Y(q_{\tau 0})(q_\tau(\cdot) - q_{\tau 0}(\cdot))\}]|_\infty M_w \\
=&|F_Y(q_{\tau 0}(\cdot)) - F_Y(q_\tau(\cdot)) + f_Y(q_{\tau 0})(q_\tau(\cdot) - q_{\tau 0}(\cdot))|_\infty M_w = o(|q_\tau(\cdot) - q_{\tau 0}(\cdot)|_\infty).
\end{aligned}
$$

Now we verify condition **G.3**. To find the pathwise derivative of $\mathrm{Z}(q_\tau, w_0)$ with respect to w, we conduct the following calculations. For any $\bar{w}$ such that $\{w_0 + \alpha(\bar{w} - w_0) : \alpha \in [0,1]\} \subset \Pi$,

$$\frac{\mathrm{E}[m(Y; q_\tau)(w_0 + \alpha(\bar{w} - w_0))] - \mathrm{E}[m(Y; q_\tau)w_0]}{\alpha} = \mathrm{E}[m(Y; q_\tau)(\bar{w} - w_0)],$$

and it has the limit $\mathrm{E}[m(Y; q_\tau)(\bar{w} - w_0)]$ as $\alpha \to 0$. Therefore $Z_2(q_\tau, w_0)[w - w_0] = \mathrm{E}[m(Y; q_\tau)(w - w_0)]$ in all directions $[w - w_0] \in \Pi$ ($Z_2(.)$ is defined in the Supplemental Appendix, **Lemma 3**). Condition **G.3.1** is satisfied by noting that

$$|\mathrm{E}[m(Y; q_\tau(\cdot))w_0(\boldsymbol{U}; \cdot)] - \mathrm{E}[m(Y; q_\tau(\cdot))w(\boldsymbol{U}; \cdot)] - \mathrm{E}[m(Y; q_\tau(\cdot))(w - w_0)(\boldsymbol{U}; \cdot)]|_\infty = 0.$$

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

$$\begin{aligned} &|\mathrm{E}[m(Y; q_\tau(\cdot))(w - w_0)(\boldsymbol{U}; \cdot)] - \mathrm{E}[m(Y; q_{\tau 0}(\cdot))(w - w_0)(\boldsymbol{U}; \cdot)]|_\infty \\ =&|\mathrm{E}[m(Y; q_\tau(\cdot)) - m(Y; q_{\tau 0}(\cdot))(w - w_0)(\boldsymbol{U}; \cdot)]|_\infty \\ \leq&|\mathrm{E}[m(Y; q_\tau(\cdot))] - \mathrm{E}[m(Y; q_{\tau 0}(\cdot))]|_\infty o(1) = \delta_n o(1), \end{aligned}$$

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

Proposition 1 establishes the results **PG.1**–**PG.3**, which are uniform results, and useful for the derivations. Condition **G.4** is automatically satisfied by **PG.3**. Now we check condition **G.5**. To do so, we check that $\{\psi_{q,w,t} : q \in \ell_\delta^\infty(\mathcal{T}), w \in \Pi, t \in \mathcal{T}\}$, where $\psi_{q,w,t} = (\tau - \mathbf{1}\{Y < q_\tau(t)\})w(\boldsymbol{U}; t)$, is Donsker, which in turn implies **G.5'**. This result follows because by result **PG.1** in Proposition 1 and the Corollary 2.7.4 in van der Vaart and Wellner (1996) the bracketing number of $\Pi = \{w \in \Pi : |w - w_0|_\infty < \delta\}$ is finite. Thus Π is Donsker with a constant envelope. The class $\mathcal{F} = \{\mathbf{1}\{Y < q_\tau(t)\} : q \in \ell_\delta^\infty(\mathcal{T})\}$ is Donsker by exploiting the monotonicity and boundedness of indicator function and bounded density condition assumed in **PC.4**. Finally, the result follows because the class $\psi_{q,w,t}$ 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) $\psi_{q,w,t}$ is Donsker and we have that **G.5'** is satisfied by Lemma 3.3.5 of an der Vaart and Wellner (1996). As noted in the Supplemental Appendix **G.5'** implies **G.5**.

Finally, condition **G.6** is implied by the result in **PG.2**, which is derived in Proposition 1.

Hence, all the conditions of **Lemma 3** are satisfied.

■

Proof of Corollary 3. The result follows from Theorem 2, which establishes weak convergence for each given t, and the continuous mapping theorem. ■

Proof of Corollary 4. The assertion holds by **Theorem 2** and the continuous mapping theorem. ■

Proof of Corollary 5. The assertion follows by the definition of $\Delta_\tau(t + \delta, t)$, **Theorem 2** and the continuous mapping theorem. ■

Proof of Proposition 2. The proof is a direct application of **Lemma 4** in the Online Supplemental Appendix. ■

10. Supplemental Appendix (Online)

This supplement presents results for the asymptotic theory for the generic Z-estimator. The results are given in Galvao and Wang (2015) and we reproduce them for easy of exposition.

10.1. Asymptotic Theory

This appendix establishes some asymptotic properties of a generic Z-estimator. Lemmas 2 and 3 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. Lemma 4 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.

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 $\mathbb{Z}_n : \Theta \times \mathcal{H} \mapsto \mathcal{L}$, $\mathrm{Z} : \Theta \times \mathcal{H} \mapsto \mathcal{L}$ be random maps and a deterministic map, respectively. We suppress the dependence of $\mathbb{Z}$ on n for simplicity. The Z-estimator $\widehat{\theta}$ is defined as the approximate root of

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

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

10.2. 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 $||\mathrm{Z}(\theta_n, h_0)||_{\mathcal{L}} \to 0$ implies $\theta_n \to \theta_0$ for all sequences $\theta_n \in \Theta$.

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

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

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

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

Condition **C.1** requires that $\widehat{\theta}$ solves the estimating equation $||\mathbb{Z}(\theta, \widehat{h})||_{\mathcal{L}} = 0$ 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 $\delta_n \downarrow 0$,

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

This set of conditions is similar to the conditions of Theorem 1 of Chen et al. (2003).

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

Lemma 2. *Suppose that* $\theta_0 \in \Theta$ *satisfies* $\mathrm{Z}(\theta_0, h_0) = 0$ *with* $h_0 \in \mathcal{H}$ *and that conditions* **C.1–C.5** *hold. Then* $||\widehat{\theta} - \theta_0||_\Theta = o_{p^*}(1)$.

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

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

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

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

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 $||\mathrm{Z}(\widehat{\theta}, h_0)||_{\mathcal{L}} \leq o_{p^*}(1)$ and hence the result. ■

10.3. Weak Convergence

Now we provide a general result of weak convergence for Z-estimators. For the proof of weak convergence, 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 et al. (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 $\mathit{\Theta}$ and $\mathbb{L}$ denote Banach spaces, and $\mathbb{H}$ a normed space. Fréchet differentiability of a map $\phi : \mathit{\Theta} \mapsto \mathbb{L}$ at $\theta \in \mathit{\Theta}$ means that there exists a continuous, linear map $\phi'_\theta : \mathit{\Theta} \mapsto \mathbb{L}$ with

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

for all sequences $\{h_n\} \subset \mathit{\Theta}$ with $||h_n|| \to 0$ and $\theta + h_n \in \mathit{\Theta}$ 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

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

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 $\theta \mapsto \mathrm{Z}(\theta, h_0)$ is Fréchet differentiable at θ_0 with a continuously invertible derivative $\mathrm{Z}_1(\theta_0, h_0)$.

G.3 For all $\theta \in \Theta_\delta$ the pathwise derivative $\mathrm{Z}_2(\theta, h_0)[h - h_0]$ of $\mathrm{Z}(\theta, h_0)$ 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 $||\mathrm{Z}(\theta, h_0) - \mathrm{Z}(\theta, h) - \mathrm{Z}_2(\theta, h_0)[h - h_0]||_{\mathcal{L}} \leq c||h - h_0||^2_{\mathcal{H}}$ for a constant $c \geq 0$.

G.3.2 $||\mathrm{Z}_2(\theta, h_0)[h - h_0] - \mathrm{Z}_2(\theta_0, h_0)[h - h_0]||_{\mathcal{L}} \leq o(1)\delta_n$.

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 $\delta_n \downarrow 0$,

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

G.6 $\sqrt{n}(\mathrm{Z}_2(\theta_0, h_0)[\widehat{h} - h_0] + (\mathbb{Z} - \mathrm{Z})(\theta_0, h_0))$ 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 et al. (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 $\delta_n \downarrow 0$,

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

Now we provide a general result for Z-estimators.

Lemma 3. *Suppose that $\theta_0 \in \Theta_\delta$ satisfies $\mathrm{Z}(\theta_0, h_0) = 0$, that $\widehat{\theta} = \theta_0 + o_{p^*}(1)$, and that conditions* **G.1**–**G.6** *hold. Then,*

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

Proof. [Proof of **Lemma 3**]

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 $\mathrm{Z}(\theta, h_0)$ implies the existence of a continuous linear map $\mathrm{Z}_1(\theta_0, f_0)$ such that

$$\frac{||\mathrm{Z}(\theta, f_0) - \mathrm{Z}(\theta_0, f_0) - \mathrm{Z}_1(\theta_0, f_0)(\theta - \theta_0)||_{\mathcal{L}}}{||\theta - \theta_0||_{\Theta}} = o(1).$$

By the triangle inequality, it follows

$$||\mathrm{Z}_1(\theta_0, h_0)(\theta - \theta_0)||_{\mathcal{L}} \leq ||\mathrm{Z}(\theta, h_0) - \mathrm{Z}(\theta_0, h_0)||_{\mathcal{L}} + o(||\theta - \theta_0||_{\Theta}).$$

Since the derivative $\mathrm{Z}_1(\theta_0, h_0)$ is continuously invertible by condition **G.2**, there exists a positive constant c such that $||\mathrm{Z}_1(\theta_0, h_0)(\theta_1 - \theta_2)||_{\mathcal{L}} \geq c||\theta_1 - \theta_2||_{\Theta}$ for every θ_1 and $\theta_2 \in \Theta_\delta$. Therefore, it follows

$$(c - o(1))||\theta - \theta_0||_{\Theta} \leq ||\mathrm{Z}(\theta, h_0) - \mathrm{Z}(\theta_0, h_0)||_{\mathcal{L}}, \tag{10}$$

and

$$(c - o_{p^*}(1))||\widehat{\theta} - \theta_0||_{\Theta} \leq ||\mathrm{Z}(\widehat{\theta}, h_0) - \mathrm{Z}(\theta_0, h_0)||_{\mathcal{L}} = ||\mathrm{Z}(\widehat{\theta}, h_0)||_{\mathcal{L}}, \tag{11}$$

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

$$||\mathrm{Z}(\widehat{\theta}, h_0) - \mathrm{Z}(\widehat{\theta}, \widehat{h})||_{\mathcal{L}} + ||\mathrm{Z}(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}) + \mathrm{Z}(\theta_0, h_0) - \mathbb{Z}(\theta_0, h_0)||_{\mathcal{L}} + O_p(n^{-1/2}). \tag{12}$$

For the first term, we have that

$$\begin{aligned}
||\mathrm{Z}(\widehat{\theta}, h_0) - \mathrm{Z}(\widehat{\theta}, \widehat{h})||_{\mathcal{L}} \leq & ||\mathrm{Z}(\widehat{\theta}, h_0) - \mathrm{Z}(\widehat{\theta}, \widehat{h}) - \mathrm{Z}_2(\widehat{\theta}, h_0)[\widehat{h} - h_0]||_{\mathcal{L}} \\
& + ||\mathrm{Z}_2(\widehat{\theta}, h_0)[\widehat{h} - h_0] - \mathrm{Z}_2(\theta_0, h_0)[\widehat{h} - h_0]||_{\mathcal{L}} + ||\mathrm{Z}_2(\theta_0, h_0)[\widehat{h} - h_0]||_{\mathcal{L}} \\
\leq & o_{p^*}\left(n^{-1/2}\right) + o_{p^*}\left(||\widehat{\theta} - \theta_0||_{\Theta}\right) + O_{p^*}\left(n^{-1/2}\right) \\
\leq & ||\mathrm{Z}(\widehat{\theta}, h_0)||_{\mathcal{L}} \times o_{p^*}(1) + O_{p^*}\left(n^{-1/2}\right),
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second one by conditions **G.3** and **G.6**, and the third by inequality (10).

As for the second term in (12), by condition **G.5**,

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

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,

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

It then follows

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

Thus, equation (12) is bounded by

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

and the right side of the equality in (11) satisfies

$$(1 - o_{p^*}(1))||\mathrm{Z}(\widehat{\theta}, h_0)||_{\mathcal{L}} \leq O_{p^*}\left(n^{-1/2}\right). \tag{13}$$

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**,

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

Therefore, it follows that

$$
\begin{aligned}
\mathrm{Z}_1(\theta_0,h_0)\sqrt{n}(\widehat{\theta}-\theta_0)+\sqrt{n}\mathrm{Z}_2(\theta_0,h_0)[\widehat{h}-h_0]=&\sqrt{n}(-\mathrm{Z}(\widehat{\theta},\widehat{h})+\mathrm{Z}(\theta_0,h_0))+o_{p^*}(1)\\
=&\sqrt{n}(\mathbb{Z}(\widehat{\theta},\widehat{h})-\mathbb{Z}(\theta_0,h_0))+o_{p^*}(1)\\
=&\sqrt{n}(\mathrm{Z}(\theta_0,h_0)-\mathbb{Z}(\theta_0,h_0))+o_{p^*}(1),
\end{aligned}
$$

and

$$
\mathrm{Z}_1(\theta_0,h_0)\sqrt{n}(\widehat{\theta}-\theta_0)=-\sqrt{n}(\mathrm{Z}_2(\theta_0,h_0)[\widehat{h}-h_0]+(\mathbb{Z}-\mathrm{Z})(\theta_0,h_0))+o_{p^*}(1)\rightsquigarrow\mathbb{G},
$$

by condition **G.6**.

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

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

■

10.4. 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 4** we provide a result for the validity of the bootstrap for general Z-estimator. It is also an extension of that in Chen et al. (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 $\mathrm{Z}\left(\theta,h\right)=\mathrm{E}m^{\dagger}\left(W_i,\theta;h\left(W_i,\theta\right)\right)$ and $\mathbb{Z}\left(\theta,h\right)=\frac{1}{n}\sum_{i=1}^{n}m^{\dagger}\left(W_i,\theta;h\left(W_i,\theta\right)\right)$, where $\{W_i\}$ is i.i.d and $m^{\dagger}\left(\cdot\right)$ is some known function. It is without loss of generality to study only the validity of bootstrap for $\sqrt{n}(\widehat{\theta}(t)-\theta_0(t))$. Let $\widehat{h}^*$ be an estimator of h_0 using resampled data. Let $\mathbb{Z}^*(\theta,h)$ denote the resampled average. The bootstrap estimator $\widehat{\theta}^*$ satisfies

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

Following Chen et al. (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 $\delta_n \downarrow 0$,

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

G.6B $\sqrt{n}(\mathrm{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] + (\mathbb{Z}^* - \mathbb{Z})(\widehat{\theta}, \widehat{h}))$ 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 4. *Suppose $\theta_0 \in \mathrm{int}(\Theta)$ and $\widehat{\theta} \overset{a.s.}{\to} \theta_0$. 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 $\mathrm{Z}_1(\theta; h)$ 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. [Proof of **Lemma 4**]

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}) - \mathrm{Z}_1(\widehat{\theta}, \widehat{h})(\widehat{\theta}^* - \widehat{\theta}) - \mathrm{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}]|| \\
=&||\mathrm{Z}(\widehat{\theta}^*, \widehat{h}^*) - \mathrm{Z}(\widehat{\theta}^*, \widehat{h}) - \mathrm{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] + \mathrm{Z}(\widehat{\theta}^*, \widehat{h}) - \mathrm{Z}(\widehat{\theta}, \widehat{h}) - \mathrm{Z}_1(\widehat{\theta}, \widehat{h})(\widehat{\theta}^* - \widehat{\theta}) \\
&+ [(\mathbb{Z}^*(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*)) - (\mathbb{Z}^*(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}))] \\
&+ [(\mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*) - \mathrm{Z}(\widehat{\theta}^*, \widehat{h}^*)) - (\mathbb{Z}(\widehat{\theta}, \widehat{h}) - \mathrm{Z}(\widehat{\theta}, \widehat{h}))] \\
&+ \mathrm{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] - \mathrm{Z}_2(\widehat{\theta}^*, \widehat{h})[\widehat{h}^* - \widehat{h}]|| \\
\le&||\mathrm{Z}(\widehat{\theta}^*, \widehat{h}^*) - \mathrm{Z}(\widehat{\theta}^*, \widehat{h}) - \mathrm{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}]|| + ||\mathrm{Z}(\widehat{\theta}^*, \widehat{h}) - \mathrm{Z}(\widehat{\theta}, \widehat{h}) - \mathrm{Z}_1(\widehat{\theta}, \widehat{h})(\widehat{\theta}^* - \widehat{\theta})|| \\
&+ ||(\mathbb{Z}^*(\widehat{\theta}^*, \widehat{h}^*) - \mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*)) - (\mathbb{Z}^*(\widehat{\theta}, \widehat{h}) - \mathbb{Z}(\widehat{\theta}, \widehat{h}))|| \\
&+ ||(\mathbb{Z}(\widehat{\theta}^*, \widehat{h}^*) - \mathrm{Z}(\widehat{\theta}^*, \widehat{h}^*)) - (\mathbb{Z}(\widehat{\theta}, \widehat{h}) - \mathrm{Z}(\widehat{\theta}, \widehat{h}))|| \\
&+ ||\mathrm{Z}_2(\widehat{\theta}, \widehat{h})[\widehat{h}^* - \widehat{h}] - \mathrm{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

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

and

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

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

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

and the result follows. ■