

A practical generalized propensity-score estimator for quantile continuous treatment effects

Javier Alejo
IECON-UdelaR
Montevideo, Uruguay
javier.alejo@ccee.edu.uy

Antonio F. Galvao
University of Arizona
Tucson, AZ
agalvao@email.arizona.edu

Gabriel Montes-Rojas
CONICET-IIEP-UBA
Ciudad Autónoma de Buenos Aires, Argentina
gabriel.montes@fce.uba.ar

Abstract. In this article, we present a new command, `qcte`, that implements several methods for estimation and inference for quantile treatment-effects models with a continuous treatment. We propose a semiparametric two-step estimator, where the first step is based on a flexible Box–Cox model, as the default model of the command. We develop practical statistical inference procedures using bootstrap. We implement some simulations to show that the proposed methods perform well. Finally, we apply `qcte` to a survey of Massachusetts lottery winners to estimate the unconditional quantile effects of the prize amount, as a proxy of nonlabor income changes, on subsequent labor earnings from U.S. Social Security records. The empirical results reveal strong heterogeneity across unconditional quantiles.

Keywords: `st00!!`, `qcte`, continuous treatment, quantile treatment effects, quantile regression

1 Introduction

The effect of policy variables on distributional outcomes are of fundamental interest in empirical economics and 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. On the binary TE models, Hahn (1998); Heckman et al. (1998); Hirano, Imbens, and Ridder (2003); Abadie and Imbens (2006); and Li, Racine, and Wooldridge (2009) study efficient estimation of the average TE. The Stata `teffects` command can be used to implement these. There is also a lot of literature on exploring effects of heterogeneity using unconditional quantiles, quantile TE. See, for example, Chernozhukov and Hansen (2005) and Firpo (2007). Many of these can be implemented following Frölich and Melly's (2010) package, `ivqte`. See also Hasebe (2013) for an alternative estimator.

There is also literature on estimation of multivalued TE, which can be implemented with the package `poparms`; see, for example, Imbens (2000); Lechner (2001); Cattaneo (2010); and Cattaneo, Drukker, and Holland (2013). It is known, however, that categorizing or discretizing continuous treatments generally leads to several serious problems, such as loss of power in testing, misclassification (which is associated with potential bias), problems with prediction, and even problems with interpretation of the results and coefficients of interest. See, for example, Cox (1957); Cohen (1983); van Belle (2008); and Fedorov, Mannino, and Zhang (2009) for more comprehensive discussions on problems associated with discretizing continuous variables. Recently, there has been a growing interest in continuous TE. Continuous treatments (such as those indexed by dose, exposure, duration, or frequency) arise often in practice, especially in observational studies. More importantly, such treatments lead to effects that are naturally described by curves (for example, dose–response curves as functionals of the treatment dose) rather than scalars (for example, point estimators) as in discrete treatments. Many articles in the literature on unconditional TE concentrate on discrete treatments, that is, binary or multivalued treatment assignments. Among others, Hirano and Imbens (2004) and Imai and van Dyk (2004) develop a generalized propensity score for continuous average treatment models to estimate average dose–response functions (ADRF) and average continuous TE (ACTE). Other estimators were proposed by Flores (2007) and Flores et al. (2012). Bia and Mattei (2008) and Bia et al. (2014) propose two commands, `gpscore` and `drf`, that compute the ADRF and ACTE using parametric and semiparametric techniques.

Galvao and Wang (2015) and Alejo, Galvao, and Montes-Rojas (2018) derive a two-step estimator for practical estimation and inference for quantile TE with continuous treatment. A parameter of interest in the presence of continuous treatment is the entire curve of quantile potential outcomes or quantile dose–response functions (QDRF). The QDRF 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 at given levels of treatment. Identification of the parameters of interest is based on the ignorability or weak unconfoundedness assumption (see, for example, Rubin [1977], Heckman et al. [1998], or Dehejia and Wahba [1999]), applying the methodology of Galvao and Wang (2015). The estimators are implemented as two-step estimators. In the first step, one estimates a ratio of conditional densities. In the second step, one performs a simple weighted quantile regression estimation where the weights are given by the ratio of conditional density functions. Alejo, Galvao, and Montes-Rojas (2018) propose a flexible Box–Cox density estimation procedure. This approach has important advantages. The first advantage is that the first step of the Box–Cox procedure is simple to implement in practice. The second advantage is that the Box–Cox procedure allows for many covariates and satisfies the required convergence rates for the first step. The Box–Cox procedure is thus flexible to accommodate empirical settings where the ignorability assumption is valid only after conditioning on a rich (possibly large) set of covariates. The numerical simulations show that the Box–Cox procedure is a flexible procedure that correctly estimates QDRF and QCTE functions for alternative data-generating processes.

In this article, we present a new command, `qcte`, that estimates both QDRF and QCTE. Unconditional quantile heterogeneity thus complements the results of ADRF and ACTE.

We evaluate the finite-sample performance of the proposed estimator in two ways. First, we implement Monte Carlo simulation exercises. In particular, we evaluate location shift and scale-location shift data-generating processes. Second, to illustrate the methods, we estimate the effects of nonlabor income changes on labor earnings. We use the survey of Massachusetts lottery winners and estimate the effect of the prize amount, as a proxy of exogenous nonlabor income changes, on subsequent labor earnings (from U.S. Social Security records). This database was originally used by Imbens, Rubin, and Sacerdote (2001) and then by Hirano and Imbens (2004). 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 (that is, 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. The quantile process shows 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. These results are important for analyzing the effect of general income transfers as conditional cash transfer programs in developing countries because 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 article is organized as follows. In section 2, we review the QCTE estimator of Alejo, Galvao, and Montes-Rojas (2018). In section 3, we describe the `qcte` syntax. In section 4, we illustrate the procedure by applying the command to Monte Carlo simulation and to the empirical application of the survey of Massachusetts lottery winners used by Hirano and Imbens (2004). In section 5, we conclude with practical suggestions on the proper use of the command.

2 Continuous treatment effects

We want 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, $\mathcal{T} = \{0, 1\}$. For practical convenience, we assume \mathcal{T} to be an interval $[t_0, t_1]$, and therefore the dose values of interest will be on that compact set.

An important parameter of interest when the treatment is continuous is the QDRF, which is defined as

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

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 QCTE, which is defined as

$$\Delta_\tau(t, t') := q_\tau(t) - q_\tau(t') \quad (1)$$

for $t' < t$. The QCTE, as defined in (1), 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 (2018) 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 article, 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 the QCTE defined above has a causal interpretation, that is, the effect of changing the level of the treatment for any particular subpopulation. See Firpo (2007) and Cattaneo (2010) for detailed discussions on rank preservation in quantile TE and definitions of concepts. Of particular interest is analyzing the QCTE for a fixed change in the dose, say, $\delta > 0$, over the doses $t \in \mathcal{T}$ as

$$D_\tau(t, \delta) := \Delta_\tau(t + \delta, t) = q_\tau(t + \delta) - q_\tau(t) \quad (2)$$

Unfortunately, as is usual in the TE 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 is usual in economic and other nonexperimental problems, then selection biases arise because the observed outcomes might not be the result of the dose itself but of 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 \mathbf{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 \mathbf{X} to obtain an unconditional TE. By the law of iterated expectations, unconditional expectations can be recovered.

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

$$E[m\{Y(t); q_\tau(t)\}] = 0 \quad (3)$$

Thus, $q_\tau(t)$ is defined as the solution to the moment condition given by (1). If this problem has a unique solution, the identification result relies on the following equality,

$$E[m\{Y(t); q_\tau(t)\}] = E[m\{Y; q_\tau(t)\}w_0(\mathbf{U}; t)]$$

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

$$E[m\{Y; q_\tau(t)\}w_0(\mathbf{U}; t)] = 0 \quad (4)$$

if and only if $q_\tau(t) = q_{\tau_0}(t)$.

This result is a direct application of 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 (\mathbf{X}, Y, T) equipped with a proper estimation of the density function of the treatment conditional on (\mathbf{X}, Y) .

As in the TE literature, the identification induces an estimating equation with two pieces, the function $m(\cdot)$ with a weighting function $w_0(\cdot)$. In our case, the weights are given by $\{f_{T|\mathbf{X},Y}(t|\mathbf{x}, y)\}/\{f_{T|\mathbf{X}}(t|\mathbf{x})\}$. 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 $\{f_{Y|\mathbf{X},T}(y|\mathbf{x}, t)\}/\{f_{Y|\mathbf{X}}(y|\mathbf{x})\}$. 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 articles that followed, where they estimate only $f_{Y|\mathbf{X}}(y|\mathbf{x})$, the so-called generalized propensity score. However, Hirano and Imbens’s approach also requires one to estimate $E[Y|X, T]$, or in fact, $E[Y|f_{T|\mathbf{X}}(t|x), T]$. Thus, Hirano and Imbens’s procedure and ours involve two different functional estimates to compute the parameter of interest.

Finally, because the QCTE is the difference between the QDRF at two different treatment doses, identification of QCTE, $\Delta_\tau(t, t')$, is as straightforward as the previous result.

2.1 Two-step estimator

Using the identification expression (4), Alejo, Galvao, and Montes-Rojas (2018) propose a two-step estimator for both QDRF and QCTE, in (1) and (2), respectively, as in Firpo (2007), Cattaneo (2010), and Galvao and Wang (2015). In the first step, one estimates the weights, that is, the ratio of densities, $w(\mathbf{u}; t) := \{f_{T|\mathbf{X},Y}(t|\mathbf{x}, y)\}/\{f_{T|\mathbf{X}}(t|\mathbf{x})\}$. The second step is given by a reweighted 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 (\mathbf{X}_i, Y_i, T_i) , indexed by $i = 1, \dots, n$. For each unit i , \mathbf{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 \mathbf{X}_i , the treatment received T_i , and the observed outcome corresponding to the level of the treatment received, Y_i .

First step: Estimation of w_0

To implement the estimator, we need an estimator for w_0 . In practice, one estimates $f_{T|\mathbf{X},y}(t|\mathbf{x}, y)$ and $f_{T|\mathbf{X}}(t|\mathbf{x})$ 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. The most important is that in several empirical applications, the number of variables in \mathbf{X} is relatively large, and because it is well known in the literature, it has an adverse effect on nonparametric methods because of the curse of dimensionality. Therefore, there are compelling reasons to use flexible parametric models to estimate the ratio of the conditional density functions. Following the results of Carroll and Ruppert (1984), Alejo, Galvao, and Montes-Rojas (2018) suggest a flexible Box–Cox estimation. This approach has important advantages. The first advantage is that the first step of the Box–Cox procedure is quick and simple to implement in practice. The second advantage is that the Box–Cox procedure allows for many covariates and satisfies the required convergence rates for the first step.

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

$$\Lambda(T, \lambda_1) = \Lambda(\{\mathbf{X}\}, \lambda_2)\boldsymbol{\beta}_X + \Lambda(Y, \lambda_2)\beta_Y + \epsilon$$

where $\epsilon|\mathbf{X}, Y \sim N(0, \sigma_\epsilon^2)$ and $\Lambda(\cdot, \lambda)$ is the Box–Cox transformation function, which is defined as $\Lambda(Z, \lambda) = \log Z$ if $\lambda = 0$ and $= (Z^\lambda - 1)/\lambda$ otherwise. Using maximum likelihood estimation, we obtain the unknown set of parameters $\boldsymbol{\mu} := (\lambda_1, \lambda_2, \boldsymbol{\beta}_X, \beta_Y, \sigma_\epsilon^2)$ and finally the conditional density $\hat{f}_{T|\mathbf{X},Y}(t|\mathbf{x}, y)$ (see appendix A.1 for more details on density formulas). Moreover, we obtain $\hat{f}_{T|\mathbf{X}}(t|\mathbf{x})$ similarly using the Box–Cox model

$$\Lambda(T, \lambda_1) = \Lambda(\{\mathbf{X}\}, \lambda_2)\boldsymbol{\beta}_X + \epsilon$$

The Box–Cox transformation applies only 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. Note that the normality assumption is a simplifying condition. The Monte Carlo simulations in Alejo, Galvao, and Montes-Rojas (2018) show that the Box–Cox Gaussian model performs well for a large family of distributions.

As noted by an anonymous referee, the QDRF and QCTE models rely on the assumption that the true conditional density $f_{T|\mathbf{X},Y}(t|\mathbf{x}, y)$ or $f_{T|\mathbf{X}}(t|\mathbf{x})$ can be consistently estimated with the Box–Cox procedure. Because we do not know the structure of the conditional models, goodness-of-fit procedures can be used to evaluate the adequacy of the proposed models. In particular, because the Box–Cox estimator is maximum likelihood, the likelihood-ratio test or Akaike and Bayesian criteria can be used to select the best model.

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

Following (3), the identification condition for $q_{\tau 0}(t)$ is $\mathbf{E}([\tau - \mathbf{1}\{Y < q_{\tau 0}(t)\}]w_0(\mathbf{U}; t)) = 0$. Thus, an estimator for the QDRF $q_{\tau 0}(t)$ is

$$\hat{q}_{\tau}(t) = \arg \min_q \frac{1}{n} \sum_{i=1}^n \hat{w}(\mathbf{u}_i; t) \rho_{\tau}(y_i - q) \quad (5)$$

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 (\mathbf{X}, T, Y) , one first computes \hat{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 \hat{w} as weights as given in (5), for each given t taken over a discretized subset (that is, grid) of \mathcal{T} .

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

$$\hat{\Delta}_{\tau}(t, t') = \hat{q}_{\tau}(t) - \hat{q}_{\tau}(t')$$

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

2.2 Inference procedures

Alejo, Galvao, and Montes-Rojas (2018) show uniform consistency and weak convergence of this two-step estimator. In this section, we turn our attention to inference on both the QDRF and QCTE. First, for testing QDRF, we consider the 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 $r \in \ell^{\infty}(\mathcal{T})$. Inference is carried out uniformly over the set of treatment levels, \mathcal{T} . The basic inference process is

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

General hypotheses on the vector $q_{\tau}(t)$ can be accommodated through functions of $Q_n(\cdot)$. We consider the Kolmogorov–Smirnov and Cramér–von Mises 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 many tests using variants of the proposed tests. Note that inference for a single point estimation for a fixed level of treatment can be seen as a particular case of uniform inference with $r(t) = q_0$ and $\mathcal{T} = t$. Alejo, Galvao, and Montes-Rojas (2018) show that simple hypotheses testing for fixed t can be based on Wald statistics.

For uniform inference for QCTE, we consider general null hypothesis

$$H_0 : \Delta_{\tau_0}(t, t + \delta) - s(t) = 0 \quad t \in \mathcal{T}$$

uniformly, where δ is a fixed treatment increment, $s(t)$ 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

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

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

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

Note that point inference for two different treatment values, say, t and t' , can be stated as a particular case with $\delta = t' - t$, $r(t) = \Delta_0$, and $\mathcal{T} = t$. Again, the Wald statistic is valid in this particular case.

In practice, the procedure is implemented in a discretized subset, most conveniently on intervals of equal size, $\mathcal{T} = [t_1, \dots, t_m]$, $t_1 < \dots < t_m$. The weak limits of T_{1n} , T_{2n} , T_{3n} , and T_{4n} are functionals of Gaussian processes, and the estimation of their covariance kernel is difficult to compute. Therefore, to make practical inference, Alejo, Galvao, and Montes-Rojas (2018) suggest using simple bootstrap techniques to approximate the limiting distribution.

3 The qcte command

3.1 Syntax

The command syntax is

```
qcte depvar treatvar [if] [in] [, xvar(varlist) zvar(varlist) t0(real) t1(real)
    dt(real) quantile(#) ynotrans reps(#) nograph notest]
```

3.2 Options

`xvar(varlist)` specifies the transformed control variables (Box–Cox model).

`zvar(varlist)` specifies the nontransformed control variables.

`t0(real)` sets the value of t_0 . The default is the first percentile of T .

`t1(real)` sets the value of t_1 . The default is the last percentile of T .

`dt(real)` sets the value of δ . The default is $(t_1 - t_0)/19$.

`quantile(#)` specifies the quantile to estimate. The default is `quantile(50)`.

`ynotrans` specifies to not transform the dependent variable.

`reps(#)` specifies the number of bootstrap replications to be performed. The default is `reps(50)`.

`nograph` suppresses the display of QDRF and QCTE plots.

`notest` suppresses the table display of statistics for uniform inference.

3.3 Stored results

`qcte` stores the following in `r()`:

Scalars

<code>r(LRx)</code>	likelihood ratio of Box–Cox model for T as a function of X
<code>r(AICx)</code>	Akaike information criterion of Box–Cox model for T as a function of X
<code>r(BICx)</code>	Bayesian information criterion of Box–Cox model for T as a function of X
<code>r(LRxy)</code>	likelihood ratio of Box–Cox model for T as a function of X and Y
<code>r(AICxy)</code>	Akaike information criterion of Box–Cox model for T as a function of X and Y
<code>r(BICxy)</code>	Bayesian information criterion of Box–Cox model for T as a function of X and Y

Matrices

<code>r(QDRFplot)</code>	matrix with numerical coordinates of the QDRF plot and their confidence intervals
<code>r(QCTEplot)</code>	matrix with numerical coordinates of the QCTE plot and their confidence intervals
<code>r(UITest)</code>	matrix with the results of the uniform inference tests

Matrices `r(QDRFplot)` and `r(QCTEplot)` are useful to replicate the output plot with other graph formats. `r(UITest)` stores the statistics T_{1n} , T_{2n} , T_{3n} , and T_{4n} , the critical values, and the p -values computed by bootstrap using $r(t) = 0$ and $s(t) = 0$.

4 Examples

In this section, we present the syntax of the `qcte` command, which implements the methodology suggested by Alejo, Galvao, and Montes-Rojas (2018) using two examples. First, we show some exercises with simulated data to show the basic output of the command on the screen. Second, we use `qcte` with real data using the base of winners of the lottery of Massachusetts.

4.1 Example 1: Simulations

For comparison, we develop some examples in Alejo, Galvao, and Montes-Rojas (2018) by drawing random samples from data-generating processes: $X = 20 + v_1$, $T = X + v_2$, and $Y = T + X + \{1 + \alpha(20 - t)^2\}v_3$ with v_1 , v_2 , and v_3 independent random variables. The parameter α determines whether the treatment effect is a pure location shift ($\alpha = 0$) or a scale-location shift $\alpha \neq 0$.

First, we evaluate the performance for a location shift treatment effect with standard normal distributions for v_1 , v_2 , and v_3 :

```
. * Location effect
. set seed 010101
. set obs 500
number of observations (_N) was 0, now 500
. generate control = rnormal(20,1)
. generate treat = control + rnormal(0,1)
. generate result = treat + control + rnormal(0,1)
. qcte result treat, xvar(control) quantile(25) reps(200) t0(15) t1(25) dt(2)
> nograph notest
(running qcte_est on estimation sample)

Bootstrap replications (200)
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
..... 50
..... 100
..... 150
..... 200

Dose-Response Function
      treat      ADRF      QDRF25      SE [95% Conf. Interval]
r1    15.0000    35.9034    35.5822    1.2021    33.2262    37.9383
r2    17.0000    37.4851    36.0936    0.5767    34.9632    37.2240
r3    19.0000    39.0709    38.2071    0.1389    37.9348    38.4794
r4    21.0000    41.0310    40.0430    0.1602    39.7291    40.3569
r5    23.0000    43.4251    42.3897    0.6795    41.0580    43.7214
r6    25.0000    45.2178    44.4872    0.8209    42.8783    46.0961

Treatment Effect
      treat      ACTE      QCTE25      SE [95% Conf. Interval]
r1    17.0000    1.5817    0.5114    1.1948    -1.8304    2.8531
r2    19.0000    1.5858    2.1135    0.5489    1.0377    3.1893
r3    21.0000    1.9601    1.8359    0.1920    1.4596    2.2122
r4    23.0000    2.3941    2.3467    0.6497    1.0732    3.6202
r5    25.0000    1.7927    2.0976    1.0195    0.0995    4.0957

. drop _all
```

Second, we consider a random sample from a scale-location shift ($\alpha = 1/5$) of the treatment with standard normal distributions for v_1 , v_2 , and v_3 :

```
. * Scale-location effect
. set seed 010101
. set obs 500
number of observations (_N) was 0, now 500
. generate control = rnormal(20,1)
. generate treat = control + rnormal(0,1)
. generate result = treat + control + (1+0.2*(treat-20)^2)*rnormal(0,1)
```

```
. qcte result treat, xvar(control) quantile(25) reps(200) t0(15) t1(25) dt(2)
> nograph notest
(running qcte_est on estimation sample)

Bootstrap replications (200)
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
..... 50
..... 100
..... 150
..... 200

Dose-Response Function
      treat      ADRF      QDRF25      SE [95% Conf. Interval]
r1    15.0000    39.3953    38.1531    3.2391    31.8046    44.5016
r2    17.0000    40.3706    41.2223    2.0013    37.2998    45.1449
r3    19.0000    40.4440    39.8571    1.2084    37.4886    42.2255
r4    21.0000    40.7902    39.7543    0.2426    39.2788    40.2299
r5    23.0000    43.2189    41.1529    0.7866    39.6111    42.6946
r6    25.0000    48.1000    46.0331    2.0495    42.0162    50.0501

Treatment Effect
      treat      ACTE      QCTE25      SE [95% Conf. Interval]
r1    17.0000    0.9754    3.0692    2.4149    -1.6639    7.8024
r2    19.0000    0.0734    -1.3653    1.2595    -3.8338    1.1033
r3    21.0000    0.3461    -0.1027    1.0657    -2.1915    1.9861
r4    23.0000    2.4287    1.3985    0.7618    -0.0946    2.8916
r5    25.0000    4.8811    4.8803    1.9811    0.9974    8.7632

. drop _all
```

Third, we consider a scale-location shift model ($\alpha = 1/5$) with a standardized χ_3^2 for v_3 . This case is characterized by the asymmetry due to a large mass of probability on the right tail of the distribution.

```
. * Scale-location effect (chi2)
. set seed 010101
. set obs 500
number of observations (_N) was 0, now 500
. generate control = rnormal(20,1)
. generate treat = control + rnormal(0,1)
. generate result = treat + control + (1+0.2*(treat-20)^2)*(rchi2(3)-3)/sqrt(6)
. qcte result treat, xvar(control) quantile(25) reps(200) t0(15) t1(25) dt(2)
> nograph notest
(running qcte_est on estimation sample)

Bootstrap replications (200)
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
..... 50
..... 100
..... 150
..... 200

Dose-Response Function
      treat      ADRF      QDRF25      SE [95% Conf. Interval]
r1    15.0000    30.7218    29.6082    3.5341    22.6815    36.5349
r2    17.0000    38.6519    36.4843    2.9177    30.7656    42.2029
r3    19.0000    41.0636    39.1163    1.7686    35.6499    42.5827
r4    21.0000    41.2719    39.8111    0.3268    39.1706    40.4516
r5    23.0000    43.9633    43.5101    1.0506    41.4510    45.5692
r6    25.0000    45.0840    44.5534    0.8333    42.9202    46.1867
```

```

Treatment Effect
      treat      ACTE      QCTE25      SE [95% Conf. Interval]
r1    17.0000    7.9302    6.8761    3.4414    0.1310    13.6212
r2    19.0000    2.4117    2.6320    1.9486   -1.1871    6.4512
r3    21.0000    0.2083    0.6948    1.7034   -2.6438    4.0335
r4    23.0000    2.6914    3.6990    0.9776    1.7829    5.6151
r5    25.0000    1.1207    1.0433    1.0853   -1.0839    3.1704
. drop _all

```

The output shows two tables: the top with the estimates for the dose–response function and the bottom for TE. Each treatment value is shown with its standard errors and the 95% confidence intervals computed via bootstrap. Note that in the three examples, we have set a grid of values for $\mathcal{T} = \{15, 17, \dots, 25\}$. By default, the QDRF in the output table is evaluated at 20 equidistant points between the first and last percentile of T . The estimated QCTE is the difference between each of the QDRF points. For simplicity, the plots and the uniform inference statistics have been omitted using options `nograph` and `notest`, respectively. The following example shows those command options.

4.2 Example 2: Real data

We illustrate the `qcte` command using the survey of Massachusetts lottery winners to estimate the effect of the prize amount (as a proxy of nonlabor income) on subsequent labor earnings from U.S. Social Security records. The prize amount is a continuous variable, so we apply the command to measure its effect on the quantiles of the distribution of earnings. This database is described in Imbens, Rubin, and Sacerdote (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 ADRF 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%. To remove such biases, we make the weak unconfoundedness assumption that, conditional on covariates, the lottery prize is independent of the potential outcomes.

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 $s = 1, 2, \dots, 6$). 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, Rubin, and Sacerdote (2001) and Hirano and Imbens (2004).

As noted above, the correct estimation of the QDRF and QCTE models requires that the Box–Cox implementation in the first step produce consistent estimators of the conditional densities. After each model, the proposed command stores the Akaike information criterion and Bayesian information criterion values after each Box–Cox estimation. The user can use these goodness-of-fit measures as a guide to select a model specification. Appendix A.3 shows an example of model selection comparing some usual specifications with these goodness-of-fit measures. Below, we use the specification that emerges from that simple algorithm.

A feature of the data to be considered is that about half the sample has $Y = 0$ (52%, which corresponds to 47% for male and 59% for female). That is, about half the sample is not working and receives no income six years after winning the lottery. We follow Hirano and Imbens’s (2004) and Bia and Mattei’s (2008) approach, who consider that a zero value corresponds to an observed level of income and it requires no truncation analysis. We find that for low quantiles, that is, $\tau < 0.5$, $\text{QDRF}_\tau(t) = 0, \forall t \in \mathcal{T}$. Thus, we report only the QDRF for $\tau = 0.75, 0.95$:

```
. use lotterydataset12
. drop if year6==.
(35 observations deleted)
. generate agew2=agew^2
. generate yearw2=yearw^2
. set seed 123
. qcte year6 prize, quantile(75) ynotrans xvar(agew agew2 yearw yearw2)
> zvar(male ownhs owncoll tixbot workthen yearm1 yearm2 yearm3 yearm4 yearm5
> yearm6) reps(200)
(running qcte_est on estimation sample)

Bootstrap replications (200)
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 1     2     3     4     5
x...x...x...x...x...x...x...x...x...xx...xx...x...x    50
xx...x...xxxx...x...x...x...x...x...x...xx...x...x...  100
x...x...x...xxx...x...x...x...x...x...x...xxx...x...  150
xx...x...x...xx...xx...xx...x...xx...xx...xx...x...xxx...x  200

Uniform inference
      Stat   c95-value   p-value
T1n   2.86e+05  2.15e+05    0.0150
T2n   1.34e+05  1.15e+05    0.0150
T3n   1.04e+05  1.43e+05    0.3409
T4n   2.32e+04  2.87e+04    0.2500

. graph save qdrf75.gph, replace
(note: file qdrf75.gph not found)
(file qdrf75.gph saved)

. qcte year6 prize, quantile(95) ynotrans xvar(agew agew2 yearw yearw2)
> zvar(male ownhs owncoll tixbot workthen yearm1 yearm2 yearm3 yearm4 yearm5
> yearm6) reps(200)
(running qcte_est on estimation sample)

Bootstrap replications (200)
-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
 1     2     3     4     5
...x...x...x...x...x...x...x...x...x...x...x...x...x...  50
...x...x...x...xx...x...x...x...x...x...x...x...xx...  100
...xx...xx...x...xx...x...x...x...x...xx...x...x...xx...  150
...x...x...x...x...x...x...x...x...x...x...x...x...x...  200
```

```

Uniform inference
      Stat  c95-value  p-value
T1n  1.86e+05  1.65e+05  0.0069
T2n  1.11e+05  1.03e+05  0.0276
T3n  4.64e+04  9.68e+04  0.4621
T4n  1.09e+04  1.72e+04  0.8414
. graph save qdrf95.gph, replace
(note: file qdrf95.gph not found)
(file qdrf95.gph saved)
. graph combine qdrf75.gph qdrf95.gph, xsize(20) ysize(17) row(2)
. graph export QDRFrealdata.png, replace width(1000)
(file QDRFrealdata.png written in PNG format)
. erase qdrf75.gph
. erase qdrf95.gph

```

Figure 1 reports the ADRF with the QDRF for selected quantiles. The upper plot on the left corresponds to $\tau = 0.75$ QDRF estimates, and the bottom plot on the left corresponds to $\tau = 0.95$. The graph shows that $Y(t)$ is a decreasing function of t , and the quantile analysis has the same shape as the average effects. As in Imbens, Rubin, and Sacerdote (2001), the effects show a convex relationship suggesting a marginally decreasing effect of the lottery prize on labor earnings.

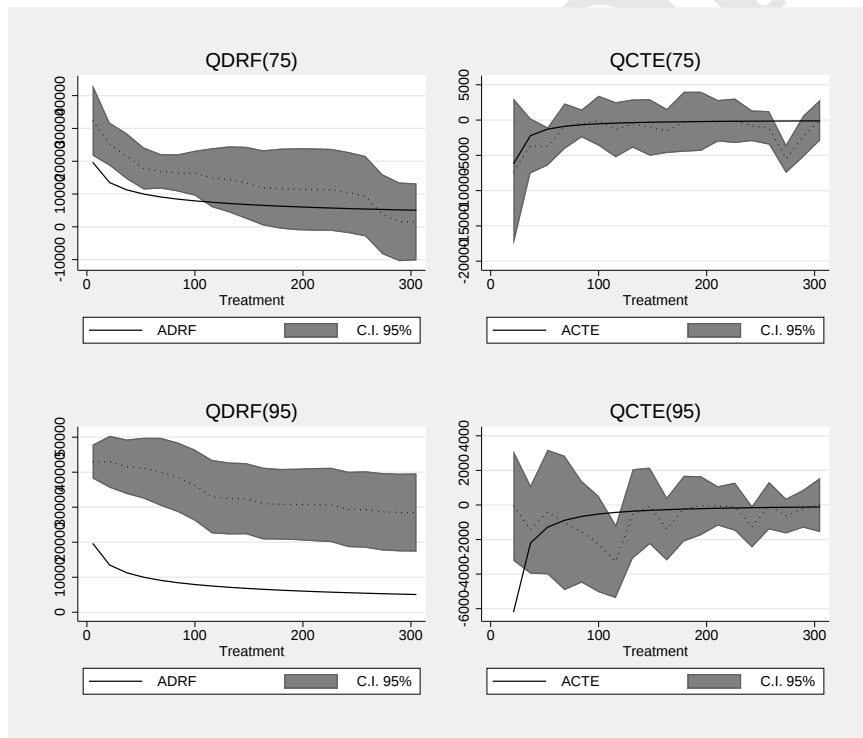


Figure 1. Empirical application: The Imbens–Rubin–Sacerdote lottery sample

Now consider inference on the point estimates and uniformly over the range of treatment values. The QDRF graph for $\tau = 0.75$ shows that the estimates are different from 0 up to a treatment value of approximately 150 (that is, for those values of t , 0 is not included in the constructed confidence interval). The QDRF for $\tau = 0.95$ and its 95% confidence interval are always above 0 for the entire treatment range. When we look at the uniform inference (T1n and T2n), the QDRF is different from zero uniformly throughout the evaluated range of treatment values for both $\tau = 0.75$ and $\tau = 0.95$. For the QCTE, the results in the graphs show that the effect of the amount won in the lottery has a nonzero treatment effect for only a few values of the continuous treatment variable. The uniform inference on QCTE (T3n and T4n), however, cannot reject the null of 0 QCTE for $\tau = 0.75$, but it rejects the null hypothesis of 0 effect at the 5% level for $\tau = 0.95$.

Finally, we compare the ADRF and ACTE estimated by our `qcte` command with those obtained using Bia and Mattei's (2008) `doseresponse` command and Bia et al.'s (2014) `drf` command. For comparison, we use the same sample (prize below its 95th quantile) and range for T as in those articles. These two commands require the previous installation of the `moremata` package (see Jann [2005] and Bia et al. [2014]). The codes used to make the comparison are shown in appendix A.2. The results are shown in figure 2. Note that `qcte` estimates are smoother than the others and that they are in between the other two. Therefore, the average effects obtained with `qcte` are consistent with the previous commands for estimating ADRF and ACTE.

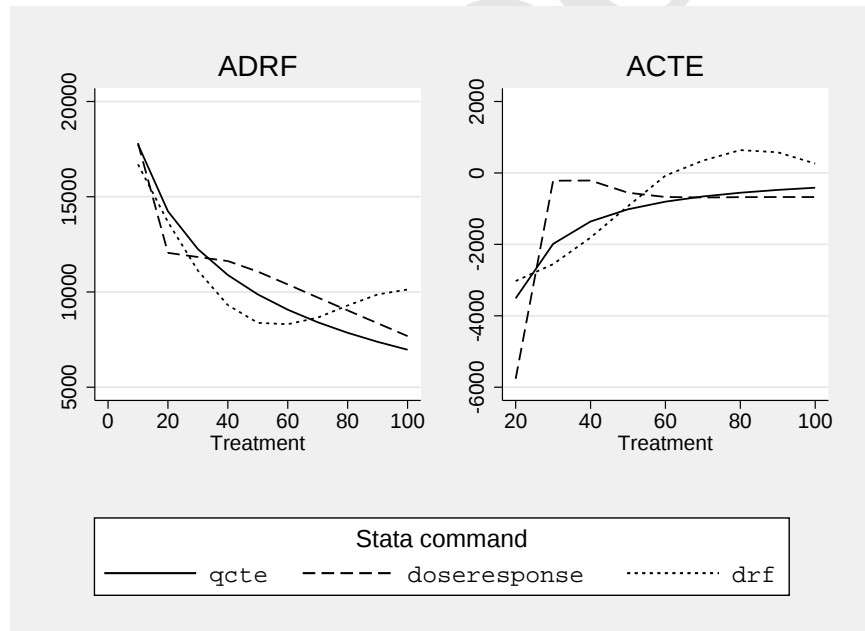


Figure 2. Command comparison: Average values

5 Conclusion

In this article, we presented a new command, `qcte`, that estimates the quantile TE models with a continuous treatment by using a semiparametric two-step estimator suggested by Galvao and Wang (2015). Following Alejo, Galvao, and Montes-Rojas (2018), we used a simple Box–Cox model to compute the propensity score and a bootstrap approach to implement these methods for many testing procedures.

Our estimates replicated the results of Alejo, Galvao, and Montes-Rojas (2018) and showed that this convexity is homogeneous in the rest of the labor earnings distribution and then showed that the threshold value was monotonic in the quantiles. The application illustrated that this method is an important tool to study continuous TE. The quantile analysis also revealed that larger prizes produce lower labor earnings, but a larger prize is required for individuals in the upper part of the distribution of unobservables. The command also provided a graphical alternative to explore heterogeneity of a continuous treatment variable.

6 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 20-2
. net install st00!! (to install program files, if available)
. net get st00!! (to install ancillary files, if available)
```

7 References

- Abadie, A., and G. W. Imbens. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74: 235–267.
- Alejo, J., A. F. Galvao, and G. Montes-Rojas. 2018. Quantile continuous treatment effects. *Econometrics and Statistics* 8: 13–36. <https://doi.org/10.1016/j.ecosta.2017.10.004>.
- Bia, M., C. A. Flores, A. Flores-Lagunes, and A. Mattei. 2014. A Stata package for the application of semiparametric estimators of dose–response functions. *Stata Journal* 14: 580–604. <https://doi.org/10.1177/1536867X1401400307>.
- Bia, M., and A. Mattei. 2008. A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score. *Stata Journal* 8: 354–373. <https://doi.org/10.1177/1536867X0800800303>.
- Carroll, R. J., and D. Ruppert. 1984. Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association* 79: 321–328. <https://doi.org/10.1080/01621459.1984.10478052>.

- Cattaneo, M. D. 2010. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics* 155: 138–154. <https://doi.org/10.1016/j.jeconom.2009.09.023>.
- Cattaneo, M. D., D. M. Drukker, and A. D. Holland. 2013. Estimation of multivalued treatment effects under conditional independence. *Stata Journal* 13: 407–450. <https://doi.org/10.1177/1536867X1301300301>.
- Chernozhukov, V., and C. Hansen. 2005. An IV model of quantile treatment effects. *Econometrica* 73: 245–261.
- Cohen, J. 1983. The cost of dichotomization. *Applied Psychological Measurement* 7: 249–253. <https://doi.org/10.1177/014662168300700301>.
- Cox, D. R. 1957. Note on grouping. *Journal of the American Statistical Association* 52: 543–547. <https://doi.org/10.2307/2281704>.
- Dehejia, R. H., and S. Wahba. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94: 1053–1062. <https://doi.org/10.1080/01621459.1999.10473858>.
- Fedorov, V., F. Mannino, and R. Zhang. 2009. Consequences of dichotomization. *Pharmaceutical Statistics* 8: 50–61. <https://doi.org/10.1002/pst.331>.
- Firpo, S. 2007. Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 75: 259–276.
- Flores, C. A. 2007. Estimation of dose–response functions and optimal doses with a continuous treatment. PhD dissertation, University of Miami. <https://core.ac.uk/download/pdf/7169663.pdf>.
- Flores, C. A., A. Flores-Lagunes, A. Gonzalez, and T. C. Neumann. 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.
- Frölich, M., and B. Melly. 2010. Estimation of quantile treatment effects with Stata. *Stata Journal* 10: 423–457. <https://doi.org/10.1177/1536867X1001000309>.
- Galvao, A. F., and L. Wang. 2015. Uniformly semiparametric efficient estimation of treatment effects with a continuous treatment. *Journal of the American Statistical Association* 110: 1528–1542. <https://doi.org/10.1080/01621459.2014.978005>.
- Hahn, J. 1998. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* 66: 315–331.
- Hasebe, T. 2013. Copula-based maximum-likelihood estimation of sample-selection models. *Stata Journal* 13: 547–573. <https://doi.org/10.1177/1536867X1301300307>.
- Heckman, J. J., H. Ichimura, J. Smith, and P. Todd. 1998. Characterizing selection bias using experimental data. *Econometrica* 66: 1017–1098.

- Hirano, K., and G. W. Imbens. 2004. The propensity score with continuous treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, ed. A. Gelman and X.-L. Meng, 73–84. Chichester, UK: Wiley.
- Hirano, K., G. W. Imbens, and G. Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71: 1161–1189.
- Imai, K., and D. A. van Dyk. 2004. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association* 99: 854–866.
- Imbens, G. W. 2000. The role of the propensity score in estimating dose–response functions. *Biometrika* 87: 706–710.
- Imbens, G. W., D. B. Rubin, and B. I. Sacerdote. 2001. Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. *American Economic Review* 91: 778–794.
- Jann, B. 2005. moremata: Stata module (Mata) to provide various functions. Statistical Software Components S455001, Department of Economics, Boston College. <https://ideas.repec.org/c/boc/bocode/s455001.html>.
- Koenker, R., and G. Bassett, Jr. 1978. Regression quantiles. *Econometrica* 46: 33–50.
- Lechner, M. 2001. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric Evaluation of Labour Market Policies*, ed. M. Lechner and F. Pfeiffer, 43–58. Heidelberg: Physica-Verlag.
- Lee, Y.-Y. 2018. Partial mean processes with generated regressors: Continuous treatment effects and nonseparable models. Department of Economics, University of California–Irvine. <https://arxiv.org/abs/1811.00157>.
- Li, Q., J. S. Racine, and J. M. Wooldridge. 2009. Efficient estimation of average treatment effects with mixed categorical and continuous data. *Journal of Business & Economic Statistics* 27: 206–223.
- Rubin, D. B. 1977. Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association* 72: 538–543.
- van Belle, G. 2008. *Statistical Rules of Thumb*. 2nd ed. Hoboken, NJ: Wiley.

About the authors

Javier Alejo is a senior researcher specializing in labor economics and inequality.

Antonio Galvao is a professor of economics specializing in econometric theory and statistics.

Gabriel Montes-Rojas is a professor of economics specializing in econometric theory and development economics.

Appendix

A.1: Density formulas

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 parameters for the dependent variable and the independent variables. Thus, we implement the estimators

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

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

where $(\hat{\lambda}_{1X}, \hat{\lambda}_{2X}, \hat{\beta}_{0X}, \hat{\beta}_{1X}, \hat{\sigma}_{T|X}^2)$ and $(\hat{\lambda}_{1YX}, \hat{\lambda}_{2YX}, \hat{\beta}_{0YX}, \hat{\beta}_{1YX}, \hat{\beta}_{2YX}, \hat{\sigma}_{T|X,Y}^2)$ are the corresponding Box–Cox parameter estimators for densities $f_{T|X}(t|X)$ and $f_{T|X,Y}(t|X, Y)$, respectively. We estimate the ratio of the densities as

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

We implement the second step with a weighted average for ADRF and weighted quantile for QDRF, where in both cases we use the corresponding $\hat{w}(X, Y, t)$ as weights. Finally, we compute ACTE and QCTE as the difference of the estimated unconditional average or quantile, respectively.

A.2: Comparison code

Below are the codes used to compare the commands. Only the point estimates are compared, so we use 25 bootstrap replications to give speed when running the algorithm. The outputs have been omitted for a better reading of the code.

```
. summarize prize, detail
  (output omitted)
. drop if prize > r(p95)
(10 observations deleted)
. generate cut = 23 if prize<=23
(126 missing values generated)
. replace cut = 80 if prize>23 & prize<=80
(90 real changes made)
. replace cut = 485 if prize>80
(36 real changes made)
```

```

. ** Alejo, Galvao, and Montes Rojas (2018)
. qcte year6 prize, quantile(75) t0(10) t1(100) dt(10) ynotrans xvar(agem agew2
> yearw yearw2) zvar(male ownhs owncoll tixbot workthen yearm1 yearm2 yearm3
> yearm4 yearm5 yearm6) reps(25)
(running qcte_est on estimation sample)
(output omitted)
. matrix PLOT = r(QDRFplot)
. ** Bia and Matei (2008)
. matrix define tp = (10\20\30\40\50\60\70\80\90\100)
. doseresponse agew ownhs male tixbot owncoll workthen yearw yearm1 yearm2
> yearm3 yearm4 yearm5 yearm6, outcome(year6) t(prize) gpscore(pscore)
> predict(hat_treat) sigma(sd) cutpoints(cut) index(p50) nq_gps(5)
> t_transf(ln) dose_response(dose_response) tpoints(tp) delta(1)
> reg_type_t(quadratic) reg_type_gps(quadratic) interaction(1)
> bootstrap(yes) boot_reps(25) filename("output-Bia2008")
(output omitted)
. ** Bia et al. (2014)
. bootstrap _b, reps(25): drf agew ownhs owncoll male tixbot workthen
> yearm1 yearm2 yearm3 yearm4 yearm5 yearm6, outcome(year6) treatment(prize)
> test(L_like) tpoints(tp) numoverlap(3) method(radialpspline)
> family(gaussian) link(log) nolog(1) search nknots(10) det delta(1)
(running drf on estimation sample)
(output omitted)
. matrix drfBia2014 = e(b)
. matrix cteBia2014 = drfBia2014[1,11..20]`
. matrix drfBia2014 = drfBia2014[1,1..10]`

```

A.3: Model-selection example

```

. cap program drop ics
. program define ics
1. scalar LR = e(chi2)
2. scalar AIC = 2*e(df_m) -2*e(l1)
3. scalar BIC = ln(e(N))*e(df_m)-2*e(l1)
4. display "Akaike's information criterion and Bayesian information
> criterion"
5. scalar list LR AIC BIC
6. scalar drop LR AIC BIC
7. end
. use lotterydataset12, clear
. generate agew2=agew^2
. generate yearw2=yearw^2
. * Model for T|X
. quietly boxcox prize, notrans(male ownhs owncoll tixbot workthen yearm1 yearm2
> yearm3 yearm4 yearm5 yearm6)
. ics
Akaike's information criterion and Bayesian information criterion
LR = 25.07667
AIC = 2339.4307
BIC = 2387.9835

```

```
. quietly boxcox prize agew, notrans(male ownhs owncoll tixbot workthen yearm1
> yearm2 yearm3 yearm4 yearm5 yearm6)
. ics
Akaike's information criterion and Bayesian information criterion
      LR = 34.743097
      AIC = 2331.7642
      BIC = 2383.7852
. quietly boxcox prize yearw, notrans(male ownhs owncoll tixbot workthen yearm1
> yearm2 yearm3 yearm4 yearm5 yearm6)
. ics
Akaike's information criterion and Bayesian information criterion
      LR = 25.100804
      AIC = 2341.4065
      BIC = 2393.4274
. quietly boxcox prize agew yearw, notrans(male ownhs owncoll tixbot workthen
> yearm1 yearm2 yearm3 yearm4 yearm5 yearm6)
. ics
Akaike's information criterion and Bayesian information criterion
      LR = 34.747187
      AIC = 2333.7602
      BIC = 2389.2491
. quietly boxcox prize agew agew2 yearw, notrans(male ownhs owncoll tixbot
> workthen yearm1 yearm2 yearm3 yearm4 yearm5 yearm6)
. ics
Akaike's information criterion and Bayesian information criterion
      LR = 37.738754
      AIC = 2332.7686
      BIC = 2391.7256
. quietly boxcox prize agew yearw yearw2, notrans(male ownhs owncoll tixbot
> workthen yearm1 yearm2 yearm3 yearm4 yearm5 yearm6)
. ics
Akaike's information criterion and Bayesian information criterion
      LR = 38.171793
      AIC = 2332.3356
      BIC = 2391.2926
. quietly boxcox prize agew agew2 yearw yearw2, notrans(male ownhs owncoll
> tixbot workthen yearm1 yearm2 yearm3 yearm4 yearm5 yearm6)
. ics
Akaike's information criterion and Bayesian information criterion
      LR = 40.820478
      AIC = 2331.6869
      BIC = 2394.112
```