ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

LIII Reunión Anual

Noviembre de 2018

ISSN 1852-0022
ISBN 978-987-28590-6-0

A practical generalized propensity score estimator for quantile continuous treatment effects

Alejo Javier
Galvao Antonio
Montes Rojas Gabriel

A practical generalized propensity score estimator for quantile continuous treatment effects

Javier Alejo
CEDLAS-UNLP and IECON-UdelaR
Montevideo/Uruguay
javieralejo@gmail.com

Antonio F. Galvao
University of Arizona
Tucson, Arizona/United States
agalvao@email.arizona.edu

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

Draft version: August 31, 2018

Abstract. A new Stata command, `qcte`, is proposed to implement several methods for estimation and inference for quantile treatment effects models with a continuous treatment. An easy semiparametric two-step estimator, where the first step is based on a flexible Box-Cox model is proposed as the default model of the command. Practical statistical inference procedures are developed using bootstrap. We implement some simulations exercises to show that the proposed methods have good performance. Finally, the command is 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.

Keywords: st0001, 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 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. Recently, there has been a growing interest on continuous TE (CTE). 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) and Li et al. (2009) study efficient estimation of the average treatment effect (ATE). There is also literature on estimation of quantile treatment effect (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.

Bia and Mattei (2008) and Bia et al. (2014) propose two Stata commands, `gpscore` and `drf`, to compute the average dose-response functions (ADRF) using parametric and semiparametric techniques. This paper develop a new Stata command, `qcte`, for a 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 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.

Identification of the parameters of interest is based on the ignorability or weak unconfoundedness assumption applying the methodology of Galvao and Wang (2015). The ignorability assumption states that given a set of observed covariates, the treatment is randomly assigned. 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 empirical estimators are implemented as two-step estimators. 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. Alejo et al. (2018) derive the asymptotic properties of the two-step estimator and develop statistical inference procedures for uniform inference and for fixed treatment values of interest. Galvao and Wang (2015) suggest a nonparametric estimation for the first step. However, there are issues with its practical implementation. First, 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. Alejo et al. (2018) propose a flexible Box-Cox density estimation procedure. This 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. The numerical simulations show 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 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 was originally used by Imbens et al. (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 (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. 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, 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 reviews the results of Alejo et al. (2018) the two-step estimator and develop statistical inference procedures. Section 3 describes the `qcte` syntax. Next we illustrate the procedure by applying the command to the the survey of Massachusetts lottery winners used by Hirano and Imbens (2004). Last, we conclude with practical suggestions on the proper use of the command.

2 Continuous Treatment Effects

The target 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

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

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.

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

$$\mathrm{E}[m(Y(t); q_\tau(t))] = 0. \tag{4}$$

thus, $q_\tau(t)$ is defined as the solution to the moment condition given by the equation (2). If this problem has a unique solution, the identification result relies on the following equality:

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

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,

$$\mathrm{E}\left[m(Y; q_\tau(t)) w_0(\boldsymbol{U}; t)\right] = 0 \tag{6}$$

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

This 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 $(\boldsymbol{X}, Y)$.

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 of the previus result.

2.1 Two step estimator

Using the identification expression (6), Alejo et al. (2018) propose a 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). 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.

First step: Estimation of w_0

To implement the estimator we need an estimator for w_0. 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. The most important is that in several empirical applications the number of variables in $\boldsymbol{X}$ is relatively large, and as it is well known in the literature, it has an adverse effect on nonparametric methods due to 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 et al. (2018) suggest a flexible Box-Cox estimation. This approach has important advantages. First, the Box-Cox first step is quick and 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

$$\Lambda(T,\lambda_1) = \Lambda((\boldsymbol{X}),\lambda_2)\beta_X + \Lambda(Y,\lambda_2)\beta_Y + \epsilon, \tag{7}$$

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 set of parameters $\mu := (\lambda_1, \lambda_2, \boldsymbol{\beta}_X, \beta_Y, \sigma_\epsilon^2)$, and finally the conditional densities $\widehat{f}_{T|\boldsymbol{X},Y}(t|\boldsymbol{x},y)$ and $\widehat{f}_{T|\boldsymbol{X}}(t|\boldsymbol{x})$.

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. It is important to highlight that the normality assumption is a simplifying condition. The Monte Carlo simulations in Alejo et al. (2018) show that the Box-Cox Gaussian model performs well for a large family of distributions.

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

Following equation (4), identification condition for $q_{\tau 0}(t)$ 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

$$\widehat{q}_\tau(t) = \arg\min_q \frac{1}{n}\sum_{i=1}^{n} \widehat{w}\left(\boldsymbol{u}_i;t\right)\rho_\tau\left(y_i - q\right), \tag{8}$$

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 (8), 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

$$\widehat{\Delta}_\tau(t,t') = \widehat{q}_\tau(t) - \widehat{q}_\tau(t'), \tag{9}$$

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

2.2 Inference procedures

Alejo et al. (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, 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 Galvao and Wang (2015) 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 $t' = t + \delta$, simple hypothesis testing can be stated as

$$H_0 : \Delta_\tau(t, t + \delta) = \Delta_0,$$

which is based on the procedures for the QDRF estimator. This, inference can be based on normality of $\sqrt{n}(\widehat{\Delta}_\tau(t, t + \delta) - \Delta_0)$. The formal justification for this procedure is based on the application of the continuous mapping theorem on the results from QDRF.

The practical implementation of the procedures will be based on the bootstrap to compute standard errors and confidence intervals based on the asymptotic normality. Also, simple hypotheses testing for fixed t can be based on Wald statistics. Consistency of this bootstrap procedure is given in Alejo et al. (2018).

3 The `qcte` syntax

3.1 Sintaxis

The command syntax is:

`qcte` *depvar treatvar* [*if*] [*in*] [`, Xvar(varlist) Zvar(varlist)`
`Quantile(#) YNOTRans Reps(#) NOGRaph`]

3.2 Options

`qcte` supports the following options:

- `xvar(varlist)` transformed control variables (Box-Cox model)
- `zvar(varlist)` do not transform specified control variables
- `ynotrans` do not transform dependent variable
- `quantile(#)` estimate #quantile; default is quantile(50)
- `reps(#)` perform #bootstrap replications; default is reps(50)
- `nograph` suppress QDRF-plot display

3.3 Saved results

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

Matrices

`r(QDRFplot)`	a matrix with numerical coordinates of the QDRF-plot and their confidence intervals.
`r(QCTEplot)`	a matrix with numerical coordinates of the QCTE-plot and their confidence intervals.

Matrices `r(QDRFplot)` and `r(QCTEplot)` are useful to replicate the output plot with other graph formats provided by Stata.

4 Examples

In this section we present the syntax of the `qcte` command that implements the methodology suggested byAlejo et al. (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 the command with real data using the base of winners of the lottery of Massachusetts.

4.1 Example 1: Simulations

For comparison purposes, we develop some examples in Alejo et al. (2018) by drawing random samples from data generating proces: $X = 20+v_1$, $T = X+v_2$ 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 if the treatment effect is a pure location shift ($\alpha = 0$) or 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

. gen control = rnormal(20,1)

. gen treat = control + rnormal(0,1)

. gen result = treat + control + rnormal(0,1)

. qcte result treat, x(control) q(25) reps(200) nograph
(running qcte_est on estimation sample)

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

```
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     QDRF_t1 |    36.1886   .6171065    58.64   0.000     34.97909    37.39811
     QDRF_t2 |   37.13925   .6304118    58.91   0.000     35.90366    38.37483
     QDRF_t3 |   37.92384   .3221879   117.71   0.000     37.29237    38.55532
     QDRF_t4 |   38.37877   .1707403   224.78   0.000     38.04412    38.71341
     QDRF_t5 |    38.9121   .2089761   186.20   0.000     38.50251    39.32168
     QDRF_t6 |   39.43871   .2837984   138.97   0.000     38.88247    39.99494
     QDRF_t7 |   40.41355   .2497981   161.78   0.000     39.92396    40.90315
     QDRF_t8 |   41.02411   .3440634   119.23   0.000     40.34976    41.69846
     QDRF_t9 |   41.78104   .4998593    83.59   0.000     40.80134    42.76075
    QDRF_t10 |   43.47282    .716724    60.65   0.000     42.06806    44.87757
     QCTE_t2 |   .9506493   .3984341     2.39   0.017     .1697329    1.731566
     QCTE_t3 |   .7845955   .3883105     2.02   0.043      .023521     1.54567
     QCTE_t4 |   .4549255   .2128516     2.14   0.033      .037744    .8721071
     QCTE_t5 |    .533329   .1242914     4.29   0.000     .2897224    .7769356
     QCTE_t6 |   .5266075   .1549067     3.40   0.001      .222996    .8302191
     QCTE_t7 |   .9748459   .1760595     5.54   0.000     .6297757    1.319916
     QCTE_t8 |   .6105614   .2296943     2.66   0.008     .1603689    1.060754
     QCTE_t9 |   .7569313   .3528742     2.15   0.032     .0653107    1.448552
    QCTE_t10 |   1.691772   .5723425     2.96   0.003     .5700018    2.813543
------------------------------------------------------------------------------
```

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
. * Scale-Location effect
. set seed 010101

. set obs 500
number of observations (_N) was 0, now 500

. gen control = rnormal(20,1)

. gen treat = control + rnormal(0,1)

. gen result = treat + control + (1+0.2*(treat-20)^2)*rnormal(0,1)

. qcte result treat, x(control) q(25) reps(200) nograph
(running qcte_est on estimation sample)

Bootstrap replications (200)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     QDRF_t1 |   41.22234   1.926301    21.40   0.000     37.44686    44.99782
     QDRF_t2 |   41.22234   1.509859    27.30   0.000     38.26307    44.18161
     QDRF_t3 |   40.07779   1.315877    30.46   0.000     37.49871    42.65686
     QDRF_t4 |   39.71772   1.138952    34.87   0.000     37.48542    41.95002
     QDRF_t5 |   39.49474   .7732594    51.08   0.000     37.97918     41.0103
     QDRF_t6 |    39.5336   .4105374    96.30   0.000     38.72896    40.33823
```

```
     QDRF_t7 |    39.8911   .2090922   190.78   0.000     39.48129    40.30092
     QDRF_t8 |    40.1626   .2554196   157.24   0.000     39.66199    40.66321
     QDRF_t9 |   40.70082   .5027129    80.96   0.000     39.71552    41.68612
    QDRF_t10 |   41.77318   1.115791    37.44   0.000     39.58627    43.96009
     QCTE_t2 |          0   .7206832     0.00   1.000    -1.412513    1.412513
     QCTE_t3 |  -1.144554     .43511    -2.63   0.009    -1.997354   -.2917541
     QCTE_t4 |  -.3600655   .3713161    -0.97   0.332    -1.087832    .3677007
     QCTE_t5 |  -.2229843     .50677    -0.44   0.660    -1.216235    .7702666
     QCTE_t6 |   .0388603   .4343697     0.09   0.929    -.8124887    .8902093
     QCTE_t7 |   .3575058   .2952852     1.21   0.226    -.2212426    .9362542
     QCTE_t8 |   .2714996    .158695     1.71   0.087    -.0395369    .5825361
     QCTE_t9 |   .5382195   .3001478     1.79   0.073    -.0500594    1.126498
    QCTE_t10 |   1.072361   .7372899     1.45   0.146    -.3727007    2.517423
------------------------------------------------------------------------------
```

Third, we consider a scale-location shift model ($\alpha = 1/5$) with a standardized χ^2_3 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

. gen control = rnormal(20,1)

. gen treat = control + rnormal(0,1)

. gen result = treat + control + (1+0.2*(treat-20)^2)*(rchi2(3)-3)/sqrt(6)

. qcte result treat, x(control) q(25) reps(200) nograph
(running qcte_est on estimation sample)

Bootstrap replications (200)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     QDRF_t1 |   36.65478   2.857324    12.83   0.000     31.05453    42.25503
     QDRF_t2 |   37.08971    2.53774    14.62   0.000     32.11583    42.06359
     QDRF_t3 |   38.75969   2.181234    17.77   0.000     34.48455    43.03483
     QDRF_t4 |   39.18689   1.568589    24.98   0.000     36.11252    42.26127
     QDRF_t5 |   39.43279   1.122366    35.13   0.000     37.23299    41.63259
     QDRF_t6 |    39.5944   .5356846    73.91   0.000     38.54448    40.64432
     QDRF_t7 |   39.94733   .2839267   140.70   0.000     39.39084    40.50381
     QDRF_t8 |   40.58221   .4651866    87.24   0.000     39.67047    41.49396
     QDRF_t9 |   41.74943   1.020816    40.90   0.000     39.74867    43.75019
    QDRF_t10 |   43.51014   .8832923    49.26   0.000     41.77891    45.24136
     QCTE_t2 |   .4349327   1.250984     0.35   0.728    -2.016952    2.886817
     QCTE_t3 |   1.669979   1.022556     1.63   0.102    -.3341948    3.674153
     QCTE_t4 |   .4272003   .8875402     0.48   0.630    -1.312346    2.166747
     QCTE_t5 |   .2458954   .6638088     0.37   0.711    -1.055146    1.546937
     QCTE_t6 |   .1616096    .694188     0.23   0.816    -1.198974    1.522193
     QCTE_t7 |   .3529282   .4415251     0.80   0.424     -.512445    1.218301
```

```
      QCTE_t8 |   .6348877   .2854299     2.22   0.026     .0754554     1.19432
      QCTE_t9 |   1.167217   .6761012     1.73   0.084    -.1579167    2.492351
     QCTE_t10 |   1.760704   .8215607     2.14   0.032     .1504747    3.370933
------------------------------------------------------------------------------
```

The output table shows the estimated QDRF and QCTE for each treatment value along with its standard errors and the 95 % confidence intervals computed via bootstrap. By default, the QDRF in the output table is evaluated at ten equidistant points between the first and last percentile of T. The estimated QCTE is the difference between each of the QDRF points.

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 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 command to measure its effect on the quantiles of the distribution of earnings. 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 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 order 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 $s = 1, 2, ..., 6$). 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).

A feature of the data to be considered is that almost half the sample has $Y = 0$ (52% which corresponds to 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. 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. We find that for low quantiles, i.e. $\tau < 0.5$, $QDRF_\tau(t) = 0, \forall t \in \mathcal{T}$. Thus, we only report the QRDF for $\tau = 0.75, 0.95$:

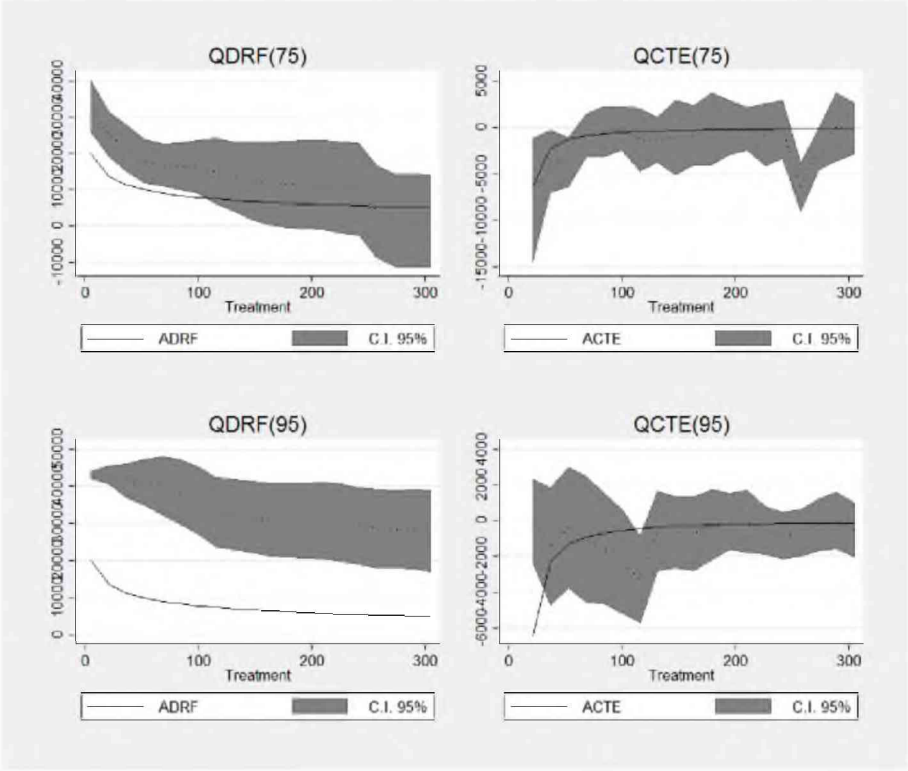
```
. use lotterydataset12.dta

.
```

```
. drop if year6==.
(35 observations deleted)

.
. qcte year6 prize, q(75) ynotrans x(agew yearw) z(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...xx.....xx.........x......    50
..........x.............x...x.........x.x....x.   100
..........xx........x...x.x...x.........xx..x...   150
.x..............x.x.x...x.........x...x..........   200
(183 missing values generated)
(20 real changes made, 1 to missing)
(file qdfr.gph saved)
(file qcte.gph saved)

. graph save qdrf75.gph, replace
(file qdrf75.gph saved)

.
. qcte year6 prize, q(95) ynotrans x(agew yearw) z(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...xx..x.x......x.x..x.x.........x.x.    50
.x......xxxx........xx.......x.....x.............   100
........x....x......xx..........x..x.......x......   150
.x.....x....x....x.x.....................xx.......   200
(183 missing values generated)
(20 real changes made, 1 to missing)
(file qdfr.gph saved)
(file qcte.gph saved)

. graph save qdrf95.gph, replace
(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)
```

Note that in this example we don't use the option `nograph` to suppress the graphic output of the command. Figure 1 reports the ADRF together with the QDRF for selected quantiles. The first plot on the left correspond to the $\tau = 0.75$ QDRF estimates, and the one on the right to $\tau = 0.95$. The graph shows an ordered cascade, where $Y(t)$ looks as a decreasing function of t. The results suggests a particular pattern in which there is a prize threshold value for which the income becomes zero.

This nonlinearity in the ADRF is explored in Imbens et al. (2001) using a quadratic specification and non-parametrically in Hirano and Imbens (2004).

Figure 1: Empirical application: the Imbens-Rubin-Sacerdote lottery sample



Both estimates shows a convex relationship suggesting a marginally decreasing effect of the lottery price on labor earnings. Our estimates replicate the results of Alejo et al. (2018) and contribute showing that this convexity is homogeneous in the rest of the labor earnings distribution and then the threshold value is monotonic in the quantiles. The application illustrates that this method is an important tool to study continuous TE. 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.

5 Coments and suggestions

This paper proposes a new Stata command `qcte` to estimate the quantile treatment effects models with a continuous treatment by using a semiparametric two-step estimator suggested by Galvao and Wang (2015). Following Alejo et al. (2018), we use a simple Box-Cox model to compute the propensity score and a bootstrap approach to implement these methods for a wide range of testing procedures. The new command `qcte` also provide a graphical alternative to explore heterogeneities of a continuous treatment variable.

6 References

Abadie, A., and G. 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* . URL `http://www.sciencedirect.com/science/article/pii/S2452306217300928`.

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

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(3): 580–604.

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.

Carroll, R. J., and D. Ruppert. 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., D. M. Drukker, and A. D. Holland. 2013. Estimation of Multivalued Treatment Effects under Conditional Independence. *Stata Journal* 13: 407–450.

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., and S. Wahba. 1999. Causal Effects in Nonexperimental Studies: Revaluating the Evaluation of Training Programs. *Journal of the American Statistical Association* 94: 1053–1062.

Fedorov, V., F. Mannino, and R. Zhang. 2009. Consequences of Dichotomization. *Pharmaceutical Statistics* 8: 50–61.

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. Mimeo.

Galvao, A., and L. Wang. 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., H. Ichimura, J. Smith, and P. Todd. 1998. Characterizing Selection Bias Using Experimental Data. *Econometrica* 66: 1017–1098.

Hirano, K., G. Imbens, and G. Ridder. 2003. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica* 71: 1161–1189.

Hirano, K., and G. W. Imbens. 2004. The Propensity Score with Continuous Treatment. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, ed. A. Gelman and X.-L. Meng. Wiley.

Imai, G., and D. van Dyk. 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., D. Rubin, and B. 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(4): 778–794.

Koenker, R., and G. W. Bassett. 1978. Regression Quantiles. *Econometrica* 46: 33–49.

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. Physical Springer, Heidelberg.

Lee, Y. 2015. Partial Mean Processes with Generated Regressors: Continuous Treatment Effects and Nonseparable Models. Mimeo.

Li, Q., J. Racine, and J. Wooldridge. 2009. Efficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous Data. *Journal of Business & Economic Statistics* 27: 206–223.

Rubin, D. 1977. Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics* 2: 1–28.

About the authors

Some background information about the first author.

Some background information about the second author.