Estimation of counterfactual distributions using quantile regression

Blaise Melly*
Swiss Institute for International Economics and Applied Economic Research (SIAW),
University of St. Gallen

April 2006

Abstract:

This paper proposes estimators of unconditional distribution functions in the presence of covariates. The conditional distribution is estimated by (parametric or nonparametric) quantile regression. In the parametric setting, we propose an extension of the Oaxaca / Blinder decomposition of means to the full distribution. In the nonparametric setting, we develop an efficient local-linear regression estimator for quantile treatment effects. We show $\sqrt{n}$ consistency and asymptotic normality of the estimators and present analytical estimators of their variance. Monte-Carlo simulations show that the procedures perform well in finite samples. An application to the black-white wage gap illustrates the usefulness of the estimators.

Keywords: Quantile Regression, Quantile Treatment Effect, Oaxaca / Blinder Decomposition, Wage Differentials, Racial Discrimination.

JEL classification: C13, C14, C21, J15, J31.

*Address for correspondence:
Blaise Melly
Bodanstrasse 8, 9000 St. Gallen, Switzerland
Phone: 0041712242301
blaise.melly@unisg.ch
http://www.siaw.unisg.ch/lechner/melly

1. Introduction

Most of the econometric literature in which the effects of a binary treatment under exogeneity are estimated has focused on average treatment effects. In the parametric setting, discrimination studies are dominated by the Oaxaca (1973) / Blinder (1973) decomposition. In the nonparametric setting, the matching literature surveyed by Imbens (2004) has focused almost entirely on the estimation of average treatment effects. Nevertheless, in many research areas, the effects of policy variables on distributional outcomes beyond simple averages are of special interest. In particular in labor economics, the distributional consequences of minimum wages, training programs and education are of primary importance to policy makers.

Motivated by this interest and by the increase in wage inequality during the last decades, studying changes in the distribution of wages has recently become an active area of research.[1] However, this literature focuses almost entirely on estimation without providing asymptotic justification or inference procedures, and it relies mostly on parametric restrictions. In this paper, we propose and derive the asymptotic distribution of a quantile equivalent of the Oaxaca / Blinder decomposition. Then, in order to relax the parametric restrictions, we propose and derive the asymptotic distribution of a local-linear-regression-based estimator for quantile treatment effects.

A regression strategy is applied in this paper. We first estimate the whole conditional distribution by (parametric and nonparametric) quantile regression. In a second step, we integrate the conditional distribution over the range of covariates in order to obtain an estimate of the unconditional distribution. The advantages of these estimators are the natural interpretability of the first step estimation and the clarity of the assumptions made. The quantile regression framework is intuitive and flexible. Due to its ability to capture heterogeneous effects, its theoretical properties have been studied extensively and it has been used in many empirical studies; see, for example, Koenker and Bassett (1978), Powell (1986), Koenker and Portnoy (1987), Chaudhuri (1991), Gutenbrunner and Jureckova (1992), Buchinsky (1994), Koenker and Xiao (2002), Angrist, Chernozhukov and Fernández-Val (2006).

This paper contributes to the existing literature in four different dimensions. First, while the basic idea of estimating the conditional distribution function by parametric quantile regression

[1] For instance, Juhn, Murphy and Pierce (1993), DiNardo, Fortin and Lemieux (1996), Gosling, Machin and Meghir (2000), Donald, Green and Paarsch (2000), Machado and Mata (2005), Lemieux (2006), Autor, Katz and Kearney (2005a and 2005b).

and integrating it to obtain the unconditional distribution is not new,[2] we propose an estimator that is faster to compute. In Section 5.2 we show that the Machado and Mata (2005) estimator, which is the most common quantile regression-based decomposition, and our proposed estimator will be numerically identical if the number of simulations used in the Machado and Mata procedure goes to infinity[3]. Hence, our asymptotic results apply also to their estimator and, since it is never possible to compute an infinite number of simulations, our estimator actually uses more information.

Second, we derive the asymptotic distribution of the parametric estimator and use the asymptotic results to propose an analytical estimator of its variance. Bootstrapping the results is time consuming and sometimes simply impossible if the number of observations is very large. The Monte-Carlo simulations show that the asymptotic results are useful approximations in finite sample. The analytical standard errors perform better than the bootstrap standard errors in our simulations.

Third, we propose a new estimator based on nonparametric quantile regression that does not require any parametric restriction. $\sqrt{n}$ consistency, asymptotic normality and achievement of the semiparametric efficiency bounds are proven. This procedure can be seen as the quantile equivalent of the estimator proposed by Heckman, Ichimura and Todd (1998) for the mean. A consistent procedure for the estimation of the variance is also presented. The estimators perform well in Monte Carlo simulations.

Finally, we apply both estimators to issues concerning racial discrimination in the USA. We first decompose the black-white wage gap using linear quantile regression. Since this parametric assumption is rejected by the data, we then use nonparametric quantile regression in the first step. The differences in basic human capital characteristics explain about one-third of the differences in the level of wages. We find that the amount of discrimination depends on the quantile at which it is evaluated but we cannot interpret the results as a glass ceiling effect.

The structure of the paper is as follows. Section 2 defines and discusses the estimands of interest. In Section 3, a parametric estimator of unconditional distributions in the presence of covariates is defined and we show how it can be used to decompose the differences in distribution. Its asymptotic distribution is then derived and an analytical estimator of its variance is proposed. Section 4 is devoted to the local-linear-regression-based matching

[2] Gosling, Machin and Meghir (2000) and Machado and Mata (2005) were the first to propose such a procedure.
[3] The Machado and Mata estimator is a simulation-based estimator.

estimator for quantile treatment effects. Section 5 presents results from different Monte-Carlo simulations. The application is presented in Section 6 and Section 7 concludes.

2. Parameters of interest and identification strategies

We are interested in the effect of a binary treatment *T* on an outcome *Y*. We have a sample of *n* units indexed by *i*, with n_0 control units and n_1 treated units. $T_i = 0$ if unit *i* receives the control treatment and $T_i = 1$ if unit *i* receives the active treatment. "Treatment" should not be taken in a restrictive sense: in the application of Section 6, $T = 0$ for whites and $T = 1$ for blacks. We use the potential-outcome notation of Neyman (1923) and characterize each unit by a pair of potential outcomes: $Y_i(0)$ for the outcome under the control treatment and $Y_i(1)$ for the outcome under the active treatment. In addition, each unit has a *K*-dimensional vector of covariates X_i. In the econometric literature, the most commonly studied estimands are the overall average treatment effect (*ATE*),

$$E\left[Y(1)\right] - E\left[Y(0)\right],$$

and the average treatment effect on the treated (*ATET*),

$$E\left[Y(1)\middle|T=1\right] - E\left[Y(0)\middle|T=1\right].$$[4]

We extend this literature by considering quantile treatment effects for the same populations, hence the overall θ^{th} quantile treatment effect (*QTE*),

$$F^{-1}_{Y(1)}(\theta) - F^{-1}_{Y(0)}(\theta),$$

and the θ^{th} quantile treatment effect on the treated (*QTET*),

$$F^{-1}_{Y(1)}\left(\theta\middle|T=1\right) - F^{-1}_{Y(0)}\left(\theta\middle|T=1\right),$$

where $F^{-1}_{Y}(\theta)$ is the θ^{th} quantile of *Y*. Note that we identify and estimate the difference between the quantiles and not the quantile of the difference. With the assumptions made in this paper we can only identify the marginal distributions of the potential outcomes but not their joint distribution. That is, we can identify the effect of a treatment on the mean, the variance, kurtosis, Gini coefficient, etc., of the distributions of the potential outcomes, but not the distribution of the individual treatment effects. In some applications, this is sufficient to answer economically meaningful questions. In welfare economics, for instance, a basic

[4] These are population measures. Imbens (2004) and Abadie and Imbens (2006) consider also the same measures conditionally on the sample. For the quantiles as for the mean effects, the only difference between the two estimands concerns the asymptotic variance and is discussed later.

assumption is anonymity. In order to compare two distributions, all permutations of personal labels are regarded as distributional equivalent (Cowell 2000) and, thus, the joint distribution is not required.

The joint distribution can be deduced from the marginal distributions if we make an additional assumption: rank invariance. This implies that the treatment does not alter the ranking of the units conditionally on *X*. This assumption is likely to be satisfied in several applications; for instance, it seems difficult to imagine that gender or race can change the ranking of an individual in the potential wage distributions. In other cases, if the rank invariance assumption is not likely to be satisfied for all observations, we can allow for given levels of overlap and bound the quantile treatment effects using the approach of Heckman, Smith and Clements (1997). In any case, knowledge of all *QTE*s is more informative than that of the *ATE*, because the mean can always be estimated by integrating over the quantiles. Since the *QTE*s have been recognized to be a useful way of summarizing the information about the distributions of the potential outcomes, we propose both estimators and inference procedures for them.

Potential outcomes are only partially observed because only $Y_i = (1-T_i)Y_i(0) + T_i Y_i(1)$ is observable. We thus need to assume that some restrictions are satisfied in order to identify the estimands of interest. In this paper, we follow the matching literature, surveyed by Imbens (2004), and assume that all regressors are exogenous. An alternative to this assumption would be the use of instrumental variables or sample selection procedures,[5] but we do not explore that approach in this paper. Our key identifying assumption is

unconfoundedness: $Y(0), Y(1) \perp T \,|\, X$.

This assumption implies, for instance, that

$$E\left[Y(0)\middle|T=1,X\right] = E\left[Y(0)\middle|T=0,X\right] = E\left[Y(0)\middle|X\right]$$

but also that

$$F^{-1}_{Y(0)}\left(\theta\middle|T=1,X\right) = F^{-1}_{Y(0)}\left(\theta\middle|T=0,X\right) = F^{-1}_{Y(0)}\left(\theta\middle|X\right).$$

When assuming unconfoundedness, parametric assumptions are a first way to identify and estimate counterfactual means and quantiles. Oaxaca (1973) and Blinder (1973) assume that the expected value of *Y* conditionally on *X* is a linear function of *X*. $E\left[Y(0)\middle|T=1\right]$ can then

[5] Abadie, Angrist and Imbens (2002), Chesher (2003), Chernozhukov and Hansen (2006), for instance, have proposed IV estimators for conditional quantile functions. Once we have obtained the coefficients corrected for endogeneity, we can use the procedure proposed in this paper to estimate quantile treatment effects (Melly 2006).

be consistently estimated by $\bar{X}^1\hat{\beta}^0_{OLS}$, where $\bar{X}^1 = n_1^{-1}\sum_{i:T_i=1} X_i$ and $\hat{\beta}^0_{OLS}$ is the vector of coefficients obtained by regressing Y on X using only control observations. They can decompose the difference between $\bar{Y}^1 = n_1^{-1}\sum_{i:T_i=1} Y_i$ and $\bar{Y}^0 = n_0^{-1}\sum_{i:T_i=0} Y_i$ into

$$\bar{Y}^1 - \bar{Y}^0 = \left[\bar{X}^1\hat{\beta}^1_{OLS} - \bar{X}^1\hat{\beta}^0_{OLS}\right] + \left[\bar{X}^1\hat{\beta}^0_{OLS} - \bar{X}^0\hat{\beta}^0_{OLS}\right].$$

The first bracket represents the effect of coefficients, typically interpreted as discrimination in numerous studies, and the second bracket gives us the effect of characteristics (justified differential). Under these assumptions, the first bracket can also be written as $E\left[Y(1)\middle|T=1\right] - E\left[Y(0)\middle|T=1\right]$ and it becomes clear that the Oaxaca / Blinder decomposition estimates the average treatment effect on the treated. If we take the treatment group as the reference, the average treatment effect on the untreated will be estimated.

In order to extend this procedure to quantiles, we need to estimate the counterfactual quantile $F^{-1}_{Y(0)}\left(\theta\middle|T=1\right)$. We assume that all quantiles of Y conditional on X are linear in X. The conditional quantiles of Y can then be estimated by linear quantile regression. Since the unconditional quantile is not the same as the integral of the conditional quantiles, we must first invert the conditional quantile function in order to obtain the conditional distribution function. Then, the unconditional distribution function can be estimated by integrating the conditional distribution function over the range of the covariates. Finally, the unconditional distribution function can be inverted in order to obtain the unconditional quantiles of interest. The details of the procedure are developed in Section 3.

For the parametric approach, we do not need to assume anything about the support of the covariates because the parametric assumption can be used to make out-of-support predictions. Obviously, one might worry about the parametric assumption, which is often arbitrary. If we want to relax the parametric restrictions, we will need to make an additional assumption: the common support condition. In order to estimate nonparametrically the counterfactual distribution of a treated unit with characteristics X, we need to find a control unit with (almost) the same characteristics. Using the notation $p(X) = \Pr\left(T=1\middle|X\right)$ and $p = \Pr(T=1)$ we can state this assumption as follows:

overlap: $0 < p(x) < 1$ for all x in the support of X.

This is the condition necessary to identify the overall quantile treatment effect. If the common support assumption is not satisfied, we can estimate the effects for the subpopulation satisfying the common support or bound the effects (Lechner 2001). For simplicity, we assume that the overlap restriction is satisfied for the whole population.

Various methods have been proposed to estimate average treatment effects assuming unconfoundedness and overlap but rejecting any parametric restriction. Following Imbens (2004), we can classify these estimators in 3 groups: matching estimators compare outcomes for pairs of observations with (almost) the same value of *X*; propensity score estimators do not adjust directly for the covariates but for the propensity score; regression methods rely on the estimation of $E\left[Y\middle|X,T=j\right]$ for $j=0,1$ and then estimate the unconditional expected value by integrating over the distribution of *X*. All strategies can be applied to estimate quantile treatment effects. Frölich (2005), for instance, follows the first approach; Firpo (2005), the second one; we follow the third approach and propose a nonparametric regression estimator for quantile treatment effect. We estimate the conditional distribution function by local-linear quantile regression (Chaudhuri 1991). Then, the unconditional distribution is obtained again by integrating the conditional distribution function over the distribution of *X*. This estimator is similar to the kernel-based estimator of Heckman, Ichimura and Todd (1998). We derive its asymptotic distribution in Section 4.

3. Parametric estimator

3.1. Model and estimators

In this section, we assume that the conditional quantiles of *Y* are linear in *X*. Extensions to general parametric assumptions are straightforward. We present an estimator of unconditional distribution functions in the presence of covariates which is then used to decompose differences in distribution, in analogy to the Oaxaca / Blinder decomposition.

Notation: $F_Y(q)$ represents the cumulative distribution of the random variable *Y* at *q*, $f_Y(q)$ represents the density of *Y* at the same point; $F_Y^{-1}(\theta)$ represents the inverse of the distribution function, commonly called the quantile function, evaluated at $0<\theta<1$; $F_Y\left(q\middle|X_i\right)$ represents the conditional cumulative distribution function of *Y* evaluated at *q* given $X=X_i$.

We make the following assumptions for $t=0,1$:

P.i. The conditional quantiles of $Y(t)$ given *X* are linear in *X*:

$$F^{-1}_{Y(t)}\left(\tau\middle|X_i\right)=X_i\beta_t\left(\tau\right),\ \text{for}\ \forall\tau\in\left(0,1\right);$$

P.ii. There exist a positive definite matrix D^0_t such that

$$\lim_{n\to\infty} n_t^{-1}\sum_{i:T_i=t} X_i'X_i=D^0_t;$$

P.iii. For $\forall\tau\in\left(0,1\right)$, there exist a positive definite matrix $D^1_t\left(\tau\right)$ such that

$$\lim_{n\to\infty} n_t^{-1}\sum_{i:T_i=t} f_{Y(t)}\left(F^{-1}_{Y(t)}\left(\tau\middle|X_i\right)\middle|X_i\right)X_i'X_i=D^1_t\left(\tau\right);$$

P.iv. For all X in the support: the distribution function $F_{Y(t)}\left(\cdot\middle|X\right)$ is absolutely continuous and has a continuous density with $0<f_{Y(t)}\left(u\middle|X\right)<\infty$ on $\left\{u:0<F_{Y(t)}\left(u\middle|X\right)<1\right\}$ and $\sup_u f'_{Y(t)}\left(u\middle|X\right)<\infty$;

P.v. $F_{Y(t)}\left(q\right)$ is absolutely continuous and has a continuous density with $0<f_{Y(t)}\left(F^{-1}_{Y(t)}\left(\theta\right)\right)<\infty$;

P.vi. For $t\in\left\{0,1\right\}$ and $t'\in\left\{0,1\right\}$, $F_{Y(t)}\left(q\middle|T=t'\right)$ is absolutely continuous and has a continuous density with $0<f_{Y(t)}\left(F^{-1}_{Y(t)}\left(\theta\middle|T=t'\right)\middle|T=t'\right)<\infty$;

P.vii. $\left\{Y_i,X_i,T_i\right\}_{i=1}^n$ are independent and identically distributed across *i* and have compact support.

Assumptions *P.i.*-*P.iv.* are traditional assumptions made in quantile regression models. Note that all assumptions are made for $\forall\tau\in\left(0,1\right)$ and $t\in\left\{0,1\right\}$ since we need to identify the whole conditional distribution of *Y* given *X* for treated and control units. Assumptions *P.v.* and *P.vi.* ensure that $F^{-1}_{Y(0)}\left(\theta\right)$, $F^{-1}_{Y(1)}\left(\theta\right)$, $F^{-1}_{Y(0)}\left(\theta\middle|T=0\right)$, $F^{-1}_{Y(0)}\left(\theta\middle|T=1\right)$, $F^{-1}_{Y(1)}\left(\theta\middle|T=0\right)$ and $F^{-1}_{Y(1)}\left(\theta\middle|T=1\right)$ are well defined and unique. They are implied by *P.iv.* if the distribution of *X* satisfies some restrictions, for instance if at least one regressor is continuously distributed on $\mathbb{R}$. To simplify the analysis and because all applications use micro-data, we assume *iid* sampling and compactness of the support.

Koenker and Bassett (1978) show that, for $t\in\left\{0,1\right\}$, $\beta_t\left(\tau\right)$ can be estimated by

$$\hat{\beta}_t\left(\tau\right)=\arg\min_{b\in\mathbb{R}^K} n_t^{-1}\sum_{i:T_i=t}\rho_\tau\left(Y_i-X_ib\right), \tag{1}$$

where ρ_τ is the *check function*

$$\rho_\tau(z) = z\left(\tau - 1(z \le 0)\right)$$

and $1(\cdot)$ is the indicator function. $\beta_t(\tau)$ is estimated separately for each τ. Asymptotically, we could estimate an infinite number of quantile regressions. In finite samples, Portnoy (1991) shows that the number of numerically different quantile regressions is $O\left(n\log(n)\right)$ and each coefficient vector prevails on an interval. Let $\left(\tau_0 = 0, \tau_1, ..., \tau_J = 1\right)$ be the points where the solution changes.[6] $\hat{\beta}_t(\tau_j)$ prevails from τ_{j-1} to τ_j for $j = 1, ..., J$.[7]

The τ's conditional quantile of $Y(t)$ given X_i is consistently estimated by $X_i'\hat{\beta}_t(\tau)$. Theoretically, it is easy to estimate the conditional distribution function by inverting the conditional quantile function. However, the estimated conditional quantile function is not necessarily monotonic and thus cannot be simply inverted. To overcome this problem, the following property of the conditional distribution function needs to be considered:

$$F_{Y(t)}\left(q|X_i\right) = \int_0^1 1\left(F_{Y(t)}^{-1}\left(\tau|X_i\right) \le q\right)d\tau = \int_0^1 1\left(X_i\beta_t(\tau) \le q\right)d\tau .$$

Thus, a natural estimator of the conditional distribution of $Y(t)$ given X_i at q is given by:

$$\hat{F}_{Y(t)}\left(q|X_i\right) = \int_0^1 1\left(X_i\hat{\beta}_t(\tau) \le q\right)d\tau = \sum_{j=1}^{J}\left(\tau_j - \tau_{j-1}\right)1\left(X_i\hat{\beta}_t(\tau_j) \le q\right). \quad (2)$$

This implies that we can estimate the unconditional distribution functions simply by

$$\hat{F}_{Y(t)}\left(q|T=t\right) = \int \hat{F}_{Y(t)}\left(q|x\right)dF_X\left(x|T=t\right) = n_t^{-1}\sum_{i:T_i=t}\hat{F}_{Y(t)}\left(q|X_i\right). \quad (3)$$

Often, we are more interested in the unconditional quantile function instead of in the unconditional distribution function since the former can be more easily interpreted.[8] Following the convention of taking the infimum of the set, a natural estimator of the θ^{th} quantile of the unconditional distribution of y is given by

$$\hat{q}_t(\theta) = \inf\left\{q : n_t^{-1}\sum_{i:T_i=t}\hat{F}_{Y(t)}\left(q|X_i\right) \ge \theta\right\}. \quad (4)$$

[6] In order to simplify the notation, we do not show the dependence of τ_j on t.

[7] We derive the results by assuming that all quantile regression coefficients have been estimated. However, the asymptotic results are also valid if we estimate quantile regression coefficients only along a grid of quantiles whose mesh is sufficiently small (a mesh size of order $O\left(n^{-1/2-\varepsilon}\right)$ will work).

[8] Juhn, Murphy and Pierce (1993), Gosling, Machin and Meghir (2000), Donald, Green and Paarsch (2000), for instance, present results for the unconditional quantile function.

Naturally, the quantiles of the unconditional distribution can be estimated consistently by the sample quantiles (Glivenko-Cantelli theorem). We will see in the next section that $\hat{q}_t(\theta)$ is more precise than the sample quantile. However, the main interest in this estimator is the possibility of simulating counterfactual quantiles that can be used to decompose differences in distribution and to estimate quantile treatment effects. For instance,

$$\hat{q}_c(\theta) = \inf\left\{ q : n_1^{-1} \sum_{i:T_i=1} \hat{F}_{Y(0)}\left(q \middle| X_i\right) \geq \theta \right\} \tag{5}$$

is the θ^{th} quantile of the distribution that we would observe if the treated units had not been treated. A decomposition of the difference between the θ^{th} quantile of the unconditional distribution of the treated and the untreated is given by:

$$\hat{q}_1(\theta) - \hat{q}_0(\theta) = \left[\hat{q}_1(\theta) - \hat{q}_c(\theta)\right] + \left[\hat{q}_c(\theta) - \hat{q}_0(\theta)\right], \tag{6}$$

where the first bracket represents the effect of coefficients (*QTET*) and the second gives us the effect of characteristics.

In the next sub-section, we concentrate on the quantiles and give the joint asymptotic distribution of $\hat{q}_1(\theta)$, $\hat{q}_0(\theta)$ and $\hat{q}_c(\theta)$, thus providing a full description of the decomposition (6). The results for quantiles of other distributions and for the estimation of other quantile treatment effects (such as the overall *QTE*) can be derived analogously. We will consider the asymptotic distribution for a single quantile in order to simplify the notation by suppressing the dependence on θ but results for the joint distribution of several quantiles are straightforward to derive.

3.2. Asymptotic results

THEOREM 1: *Under assumptions P.i. to P.vii.* $\hat{q}_0$, $\hat{q}_1$ *and* $\hat{q}_c$ *defined by (4) are consistent and asymptotically normally distributed. Define* q_0, q_1 *and* q_c *to be the true values. For* $t = 0,1$

$$\sqrt{n}\left(\hat{q}_t - q_t\right) \to N\left(0, \left(\underset{T=t}{E}\left[\left(\theta - F_{Y(t)}\left(q_t \middle| X\right)\right)^2\right] + \Omega_t \right) \Big/ \Pr\left(T=t\right) f_{Y(t)}\left(q \middle| T=t\right)^2 \right)$$

where Ω_t *is equal to*

$$\int\int f_{Y(t)}\left(q_t \middle| x\right) f_{Y(t)}\left(q_t \middle| z\right) x' \operatorname{cov}\left[\hat{\beta}_t\left(F_{Y(t)}\left(q_t \middle| x\right)\right), \hat{\beta}_t\left(F_{Y(t)}\left(q_t \middle| z\right)\right)\right] z dF_X\left(x \middle| T=t\right) dF_X\left(z \middle| T=t\right)$$

and $\operatorname{cov}\left(\hat{\beta}_t(\tau), \hat{\beta}_t(\tau')\right) = \left(\min(\tau,\tau') - \tau\tau'\right) D_t^1(\tau)^{-1} D_t^0 D_t^1(\tau)^{-1}$.

$$\sqrt{n}\left(\hat{q}_c - q_c\right) \to N\left(0, \left(\underset{T=1}{E}\left[\left(\theta - F_{Y(0)}\left(q_c|X\right)\right)^2\right] \Big/ p + \Omega_c/(1-p)\right) \Big/ f_{Y(0)}\left(q_c|T=1\right)^2\right) \qquad (7)$$

where Ω_c *is equal to*

$$\int\int f_{Y(0)}\left(q_c|x\right) f_{Y(0)}\left(q_c|z\right) x' \operatorname{cov}\left[\hat{\beta}_0\left(F_{Y(0)}\left(q_c|x\right)\right), \hat{\beta}_0\left(F_{Y(0)}\left(q_c|z\right)\right)\right] z dF_X\left(x|T=1\right) dF_X\left(z|T=1\right).$$

$\hat{q}_0$ *and* $\hat{q}_1$ *are independent. The normalized asymptotic covariance between* $\hat{q}_1$ *and* $\hat{q}_c$ *is equal to*

$$\underset{T=1}{E}\left[\left(\theta - F_{Y(1)}\left(q_1|x\right)\right)\left(\theta - F_{Y(0)}\left(q_c|x\right)\right) \Big/ f_{Y(1)}\left(q_1|T=1\right) f_{Y(0)}\left(q_c|T=1\right)\right]$$

and the normalized asymptotic covariance between $\hat{q}_0$ *and* $\hat{q}_c$ *is*

$$\frac{\int\int f_{Y(0)}\left(q_0|x\right) f_{Y(0)}\left(q_c|z\right) x' \operatorname{cov}\left[\hat{\beta}_0\left(F_{Y(0)}\left(q_0|x\right)\right), \hat{\beta}_0\left(F_{Y(0)}\left(q_c|z\right)\right)\right] z dF_X\left(x|T=0\right) dF_X\left(z|T=t\right)}{f_{Y(0)}\left(q_0|T=1\right) f_{Y(0)}\left(q_c|T=1\right)}.$$

The proof of THEOREM 1, which can be found in appendix A, is an application of theorem 2 in Chen, Linton and Van Keigelom (2003). Here, we concentrate on the interpretation of and the intuition for the results. All variances consist of two parts: the variance that we would obtain if we knew the conditional quantiles and the variance coming from the estimation of the conditional quantiles. Note that the variance of the θ^{th} sample quantile of a random variable Y can also be decomposed in this way by applying the law of total variance:

$$\begin{aligned}
&\operatorname{var}\left[1\left(Y \le q\right)\right] \Big/ f_Y\left(q\right)^2 = \left(\operatorname{var}\left[E\left(1\left(Y \le q\right)|X\right)\right] + E\left[\operatorname{var}\left(1\left(Y \le q\right)|X\right)\right]\right) \Big/ f_Y\left(q\right)^2 \\
&= \left(E\left[\left(\theta - F_Y\left(q|X\right)\right)^2\right] + E\left[F_Y\left(q|X\right)\left(1 - F_Y\left(q|X\right)\right)\right]\right) \Big/ f_Y\left(q\right)^2
\end{aligned}$$

where $F_Y\left(q\right) = \theta$. Thus, the first part of the variances of $\hat{q}_0$, $\hat{q}_1$ and $\hat{q}_c$ is the variances of the conditional quantiles. If we consider a deterministic sample or if the estimands are defined conditionally on the sample (e.g. the discussion in Imbens, 2004, and Abadie and Imbens, 2006), the variance of the estimates will consist only of the second term. In this case, uncertainty arises only from the estimation of the conditional quantile functions since the distribution of X is considered to be known. As for the estimation of the *ATE*, the variance will be lower if we estimate the sample quantity instead of the population quantity.

We also observe that the first element of the asymptotic variance of $\hat{q}_0$ and $\hat{q}_1$ is the same as the first element of the variance of the sample quantile. However, the second element is lower than for the sample quantile. The intuition is simple: the linear quantile regression model assumes that the conditional quantiles of Y given X are linear in X. All observations are used to estimate the conditional distribution function while this information does not enter the sample quantile. The price to pay is a more restrictive model. If the conditional quantile

model is misspecified, $\hat{q}_t$ is not consistent for q_t. The sample quantiles are in any case consistent. Therefore, a simple specification test of the conditional model consists of testing whether both estimates differ, like a Hausman (1978) test. If they differed significantly, it would imply that the linear quantile regression model is too restrictive.

The second part of the variances of $\hat{q}_0$, $\hat{q}_1$ and $\hat{q}_c$ is similar to the variance of the trimmed mean estimator of Koenker and Portnoy (1987) and Gutenbrunner and Jurecková (1992). The differences arise because they integrate directly over the estimated coefficients while we integrate over the estimated quantiles and because of our different assumptions concerning heteroscedasticity. An intuition for this element can be given as follows. The asymptotic variance of $\sqrt{n}X_i\hat{\beta}_t\left(\tau_j\right)$ is $X_i'\text{var}\left(\hat{\beta}_t\left(\tau_j\right)\right)X_i$. However, when estimating the θ^{th} quantile of $Y(t)$, $\hat{\beta}_t\left(\tau_j\right)$ plays only a role for those observations with $X_i\hat{\beta}_t\left(\tau_j\right)=q_t$. Moreover, the importance of each observation in estimating q_t is proportional to the density of *Y* given *X* at q_t. For instance, if the characteristics have a positive effect on *Y*, observations with a high value of *X* have a very small probability of playing a role in the estimation of a low quantile of *Y*. Finally, $\sqrt{n}X_i\hat{\beta}_t\left(F_{Y(t)}\left(q_t\middle|X_i\right)\right)$ and $\sqrt{n}X_j\hat{\beta}_t\left(F_{Y(t)}\left(q_t\middle|X_j\right)\right)$ have a covariance of $X_i'Cov\left[\hat{\beta}_t\left(F_{Y(t)}\left(q_t\middle|X_i\right)\right),\hat{\beta}_t\left(F_{Y(t)}\left(q_t\middle|X_j\right)\right)\right]X_j$ because all quantile regression coefficients are correlated. The form of the asymptotic variance-covariance for different quantile regressions was given by Gutenbrunner and Jureckova (1992).

In order to make inference on the decomposition (6), we give the covariances between $\hat{q}_c$ and $\hat{q}_0$, and between $\hat{q}_c$ and $\hat{q}_1$. They are not null because $\hat{q}_c$ and $\hat{q}_0$ are computed with the same quantile regression coefficients and $\hat{q}_c$ and $\hat{q}_1$ are computed with the same covariates, which induces co-variation of the conditional quantiles.

3.3. Estimation of the asymptotic variance

The variance of the estimators proposed in Section 3.1 can be estimated by bootstrapping the results[9]. However, since such estimators are often used with large if not huge datasets, bootstrapping the results is typically infeasible. We therefore propose to use the asymptotic results of Section 3.2 to construct an analytical estimator of the asymptotic variance.

[9] The regularity conditions for bootstrap consistency given in theorem B in Chen, Linton and Van Keigelom (2003) can be verified in the same way as the conditions for asymptotic normality which areverified in the appendix.

Consistent estimation of the asymptotic variance of $\hat{q}_c$ requires consistent estimation of p, $F_{Y(0)}\left(q_c|x\right)$, $f_{Y(0)}\left(q_c|T=1\right)$, $\text{cov}\left(\hat{\beta}_0(\tau),\hat{\beta}_0(\tau')\right)$ and $f_{Y(0)}\left(q_c|x\right)$.[10] We discuss now the estimation of each of these elements.

A natural estimator of $p=\Pr\left(T=1\right)$ is n_1/n. An estimator for $F_{Y(0)}\left(q_c|x\right)$ was given by (2). Since q_c is not known, we replace q_c by its consistent estimate $\hat{q}_c$. $f_{Y(0)}\left(q_c|T=1\right)$ is the derivative of $F_{Y(0)}\left(q_c|x\right)$. Thus, a first possibility to estimate this element is to use the idea of Siddiqui (1960):

$$\hat{f}_{Y(0)}\left(q_c(\theta)|T=1\right)=\frac{2h_n}{\hat{F}_{Y(0)}\left(\hat{q}_c(\theta+h_n)|T=1\right)-\hat{F}_{Y(0)}\left(\hat{q}_c(\theta-h_n)|T=1\right)}.$$

A second possibility is the use of a kernel estimator. Since we need to estimate the density of a counterfactual, unobserved distribution, we first simulate this distribution by estimating an important number of quantiles $\hat{q}_c(\theta_d)$ for $\{\theta_d\}_{d=1}^{D}$ taken from a uniform grid between 0 and 1.[11] We then use a normal kernel and the Silverman (1986) rule of thumb (other choice are of course also possible) and obtain:

$$\hat{f}_{Y(0)}\left(q_c(\theta)|T=1\right)=\frac{1}{nh_n}\sum_{d=1,\ldots,D}K\left(\frac{\hat{q}_c(\theta_d)-\hat{q}_c(\theta)}{h_n}\right). \qquad (8)$$

A large literature deals already with the estimation of the covariance matrix of the quantile regression coefficients[12]. In this paper, we would like to avoid the bootstrap in order to keep the computation time reasonable. Moreover, we cannot use rank-based estimators, since we need to estimate the whole covariance matrix. Finally, we want to allow for arbitrary dependence between the residuals and the regressors. Therefore, only two estimators can reasonably be used in order to estimate the variance of the quantile regression parameters: the Powell (1984) kernel estimator and the Hendricks and Koenker (1991) estimator. Normally, a disadvantage of the second estimator is that it needs more computation time because it requires the estimation of two additional quantile regressions for each quantile. But, since we have already estimated the whole quantile regression process anyway, this estimator is in our

[10] The estimation of other variances or covariances require the estimation of the same types of elements.
[11] In the Monte-Carlo simulations and in the application we set $D=10000$.
[12] See chapter 3 in Koenker (2005) for a recent survey.

case as fast as the kernel estimator. And since the Hendricks and Koenker estimator appears to be more precise in small samples, we focus on it to estimate $\widehat{\text{cov}}\left(\hat{\beta}_0(\tau),\hat{\beta}_0(\tau')\right)$ by

$$n_0\left[\sum_{i:T_i=0} X_i' X_i \hat{f}_{\varepsilon_\tau}\left(0\middle|X_i\right)\right]^{-1}\left(\min(\tau,\tau')-\tau\tau'\right)\sum_{i:T_i=0} X_i' X_i\left[\sum_{T_i=0} X_i' X_i \hat{f}_{\varepsilon_{\tau'}}\left(0\middle|X_i\right)\right]^{-1}$$

where $\hat{f}_{\varepsilon_\tau}\left(0\middle|X_i\right)=2h_n\Big/X_i\left(\hat{\beta}_0(\tau+h_n)-\hat{\beta}_0(\tau-h_n)\right)$ and h_n is a bandwidth that follows the Hall and Sheather (1988) rule.

Finally, $f_{Y(0)}\left(q_c\middle|x\right)$ is estimated in the same way by

$$\hat{f}_{Y(0)}\left(q_c\middle|x\right)=2h_n\Big/x\left(\hat{\beta}_0\left(\hat{F}^{-1}_{Y(0)}\left(\hat{q}_c\middle|x\right)+h_n\right)-\hat{\beta}_0\left(\hat{F}^{-1}_{Y(0)}\left(\hat{q}_c\middle|x\right)-h_n\right)\right).$$

Thus, we estimate the variance of $\hat{q}_c$ by

$$\left(\begin{array}{l} nn_1^{-2}\sum_{i:T_i=1}\left(\theta-\hat{\tau}_i\right)^2+ \\ nn_0^{-1}n_1^{-2}\sum_{i:T_i=1}\sum_{i:T_j=1}\hat{f}_{Y(0)}\left(\hat{q}_c\middle|X_i\right)\hat{f}_{Y(0)}\left(\hat{q}_c\middle|X_j\right)X_i'\widehat{\text{cov}}\left[\hat{\beta}_0(\hat{\tau}_i),\hat{\beta}_0(\hat{\tau}_j)\right]X_j \end{array}\right)\Bigg/\hat{f}_{Y(0)}\left(\hat{q}_c\middle|T=1\right)$$

where $\hat{\tau}_i=\hat{F}^{-1}_{Y(0)}\left(\hat{q}_c\middle|X_i\right)$ in order to alleviate the notation. The proof of consistency of this estimator follows from the consistency of the different elements of the variance, whis has already been proven in the cited papers and above, and from Slutsky and continuous mapping theorems. The proof is standard and will not be discussed here.

3.4. Extension: effects of residuals

Juhn, Murphy and Pierce (1993) and Lemieux (2006), among others, decompose the differences in distribution into three factors: coefficients, characteristics and residuals. Since there is a theoretical interest in several applications to identify these three sources of differences in distribution, we show how we can extend the decomposition of the preceding section in order to separate the effects of coefficients into the effects of median coefficients and residuals. This decomposition was developed and applied independently by Melly (2005b) and Autor, Katz and Kearney (2005a and 2005b).

We use the same framework as Juhn, Murphy and Pierce (1993) to decompose the differences in wage distributions between the treated and control units. If we take the median as a measure of central tendency of a distribution, we can write a simple wage equation for each group

$$Y_i(t)=X_i\beta_t(0.5)+u_{t,i}\qquad t=0,1.$$

We can isolate the effects of differences in characteristics, median coefficients and residuals. The effect of characteristics can be estimated similarly to Section 3.1. To separate the effect of coefficients from the effect of residuals, note that the τ^{th} quantile of the residuals distribution conditionally on X_i is consistently estimated by $X_i\left(\hat{\beta}_t(\tau)-\hat{\beta}_t(0.5)\right)$. We define $\hat{\beta}_{m1,r0}(\tau_j)=\left(\hat{\beta}_1(0.5)+\hat{\beta}_0(\tau_j)-\hat{\beta}_0(0.5)\right)$. Then, we estimate the distribution that would prevail if the median return to characteristics were the median return in the treated group but the residuals were distributed as in the control group by

$$\hat{q}_{m1,r0}(\theta)=\inf\left\{q: n_1^{-1}\sum_{i:T_i=1}\sum_{j=1}^{J}(\tau_j-\tau_{j-1})1\left(X_i'\hat{\beta}_{m1,r0}(\tau_j)\le q\right)\ge\theta\right\}.$$

Therefore, the difference between $\hat{q}_{m1,r0}(\theta)$ and $\hat{q}_c(\theta)$ is due to differences in coefficients since characteristics and residuals are kept at the same level. Finally, the difference between $\hat{q}_1(\theta)$ and $\hat{q}_{m1,r0}(\theta)$ is due to residuals.

The asymptotic distribution of this decomposition is straightforward to derive. All quantile regression coefficients estimated within the treated groups are independent from their control group analog. The covariance between different quantile regression coefficients was given in Section 3.2.

4. Semiparametric estimator

4.1. Model and estimators

The consistency of the estimators proposed above depend on the parametric assumption of the first step estimation. We have considered only linear quantile regression but nonlinear or censored quantile regression could also be used. In this case, we would have to change the form of $\text{cov}\left(\hat{\beta}_t(\tau),\hat{\beta}_t(\tau')\right)$ but the other results would still remain valid. The parametric assumption can be alleviated by using polynomial series or dummy variables. However, it is sometimes better to completely abandon parametric assumptions and to estimate the conditional quantile functions nonparametrically. We propose an estimator based on local linear quantile regression (Chaudhuri 1991). This procedure can be seen as the quantile equivalent of the estimator proposed by Heckman, Ichimura and Todd (1998) for the mean. They estimate the conditional mean function by local constant or local linear regression. Hahn (1998) computes the conditional mean function by series estimation. This is an alternative approach but we do not explore it in this paper.

In order to derive the asymptotic properties of these estimators, we make the following assumptions:

S.i. $\{Y_i, X_i, T_i\}_{i=1}^n$ are independent and identically distributed across *i* and have compact support, X is a *d*-dimensional continuously distributed variable[13] with $f_X(X_i)'$ continuously differentiable and bounded for all X_i in the support;

S.ii. For $\tau \in (0,1)$, $F_{Y(t)}^{-1}(\tau|X)$ is $\overline{p}$-smooth[14], where $\overline{p} > d$;

S.iii. For all X in the support: the distribution function $F_{Y(t)}(\cdot|X)$ is absolutely continuous and has a continuous density with $0 < f_{Y(t)}(u|X) < \infty$ on $\{u : 0 < F_{Y(t)}(u|X) < 1\}$ and $\sup_u f_{Y(t)}(u|X)' < \infty$;

S.iv. The bandwidth sequence h_n satisfies $\underset{N\to\infty}{\text{plim}} \frac{h_n}{a_n} = h_0 > 0$ for some deterministic sequence $\{a_n\}$ that satisfies $na_n^d / \log n \to \infty$ and $na_n^{2\overline{p}} \to c < \infty$ for some $c \geq 0$;

S.v. The kernel function $K(\cdot)$ is symmetric, supported on a compact set and Lipschitz continuous;

S.vi. The kernel function $K(\cdot)$ has moments of order 1 through $\overline{p}-1$ that are equal to zero.

These assumptions are in principle the same as those made by Heckman, Ichimura and Todd (1998) but some differences arise from the different estimands. Condition *S.ii.* guarantees that the conditional quantile functions are smooth enough to be estimated by local linear quantile regression. Condition *S.iii.* ensures that the conditional quantiles are well-defined and unique. Since the distribution of X is assumed to be continuous by *S.i.*, this also implies that the unconditional quantiles of Y are well-defined and unique. Undersmoothing, higher-order and compact support kernel are necessary in order to control the bias and the rate of convergence of the kernel regression estimator.

The procedure is very similar to the estimator that relies on linear quantile regression in the first step. The difference, however, is that the quantile regression coefficients depend on the point at which they are estimated. Formally, let

[13] Discrete regressors do not matter asymptotically.
[14] We call a function *p*-smooth when it is *p*-times continuously differentiable and its p^{th} derivative is Hölder continuous.

$$\hat{\beta}_t(\tau, x) = \arg\min_b n_t^{-1} \sum_{i:T_i=t} K\left(\frac{X_i - x}{h_n}\right) \rho_\tau(Y_i - X_i b)$$

be the τ th quantile regression coefficient estimated locally at *x*. We can allow the bandwidth to depend on *x* and τ . Then the procedure is similar to that of Section 3.1:

$$\hat{F}^S_{Y(t)}(q|X_i) = \int_0^1 1\left(X_i \hat{\beta}_t(\tau, X_i) \le q\right) d\tau = \sum_{j=1}^{J} (\tau_j - \tau_{j-1}) 1\left(X_i \hat{\beta}_t(\tau_j, X_i) \le q\right)$$

$$\hat{F}^S_{Y(t)}(q|T=t) = n_t^{-1} \sum_{i:T_i=t} \hat{F}^S_{Y(t)}(q|X_i) \tag{9}$$

and $$\hat{q}^S_t(\theta) = \inf\left\{q : n_t^{-1} \sum_{i:T_i=t} \hat{F}^S_{Y(t)}(q|X_i) \ge \theta\right\}. \tag{10}$$

Naturally we can estimate counterfactual quantiles by

$$\hat{q}^S_c(\theta) = \inf\left\{q : n_1^{-1} \sum_{i:T_i=1} \hat{F}^S_{Y(0)}(q|X_i) \ge \theta\right\}$$

and use them to estimate the quantile treatment effect on the treated

$$\widehat{QTET}(\theta) = \hat{q}^S_1(\theta) - \hat{q}^S_c(\theta).$$

In the same way, we estimate the overall quantile treatment effect by

$$\widehat{QTE}(\theta) = \inf\left\{q : n^{-1} \sum_{i=1}^{n} \hat{F}^S_{Y(1)}(q|X_i) \ge \theta\right\} - \inf\left\{q : n^{-1} \sum_{i=1}^{n} \hat{F}^S_{Y(0)}(q|X_i) \ge \theta\right\}.$$

4.2. Asymptotic results

THEOREM 2: *Under the assumptions S.i. to S.vi.* $\hat{q}^S_0$ *and* $\hat{q}^S_1$ *are* $\sqrt{n}$ *consistent and asymptotically equivalent to the sample quantiles:*

$$\sqrt{n}\left(\hat{q}^S_t - q_t\right) \to N\left(0, \frac{\theta(1-\theta)}{\Pr(T=t) f_{Y(t)}(q_t|T=t)^2}\right).$$

$\hat{q}^S_c$ *is* $\sqrt{n}$ *consistent and asymptotically normally distributed:*

$$\sqrt{n}\left(\hat{q}^S_c - q_c\right) \to N\left(0, \left(\frac{\underset{T=1}{E}\left[\left(\theta - F_{Y(0)}(q_c|X)\right)^2\right]}{p} + \frac{\Omega^S_c}{1-p}\right) \Big/ f_{Y(0)}(q_c|T=1)^2\right), \tag{11}$$

where $\Omega^S_c = \underset{T=1}{E}\left[F_{Y(0)}(q_c|X)\left(1 - F_{Y(0)}(q_c|X)\right) f_X(X|T=1) / f_X(X|T=0)\right]$.

$\hat{q}_0^S$ and $\hat{q}_1^S$ are independent. The normalized asymptotic covariance between $\hat{q}_0^S$ and $\hat{q}_c^S$ is

$$\frac{\underset{T=1}{E}\left[\left(\min\left(F_{Y(0)}\left(q_c|X\right),F_{Y(0)}\left(q_0|X\right)\right)-F_{Y(0)}\left(q_c|X\right)F_{Y(0)}\left(q_0|X\right)\right)\right]}{f_{Y(0)}\left(q_c|T=1\right)f_{Y(0)}\left(q_0|T=0\right)(1-p)}.$$

The normalized asymptotic covariance between $\hat{q}_c^S$ and $\hat{q}_1^S$ is

$$\frac{\underset{T=1}{E}\left[\left(F_{Y(0)}\left(q_c|X\right)-\theta\right)\left(F_{Y(1)}\left(q_1|X\right)-\theta\right)\right]}{f_{Y(0)}\left(q_c|T=1\right)f_{Y(1)}\left(q_1|T=1\right)p}.$$

Thus, $\widehat{QTET}$ and $\widehat{QTE}$ are consistent and asymptotically normally distributed:

$$\sqrt{n}\left(\widehat{QTET}-QTET\right)\to N\left(0,\text{avar}\left(\hat{q}_1^S\right)+\text{avar}\left(\hat{q}_c^S\right)-2\,\text{acov}\left(\hat{q}_1^S,\hat{q}_c^S\right)\right);^{15}$$

$$\sqrt{n}\left(\widehat{QTE}-QTE\right)\to N\left(0,\Omega_1^S+\frac{\Omega_2^S}{pf_{Y(0)}\left(F_{Y(0)}^{-1}(\theta)\right)^2}+\frac{\Omega_3^S}{(1-p)f_{Y(1)}\left(F_{Y(1)}^{-1}(\theta)\right)^2}\right),$$

where

$$\Omega_1^S=\frac{\text{var}\left[F_{Y(0)}\left(F_{Y(0)}^{-1}(\theta)|X\right)\right]}{f_{Y(0)}\left(F_{Y(0)}^{-1}(\theta)\right)^2}+\frac{\text{var}\left[F_{Y(1)}\left(F_{Y(1)}^{-1}(\theta)|X\right)\right]}{f_{Y(1)}\left(F_{Y(1)}^{-1}(\theta)|T=1\right)^2}$$
$$-2\frac{E\left[\left(F_{Y(0)}\left(F_{Y(0)}^{-1}(\theta)|X\right)-\theta\right)\left(F_{Y(1)}\left(F_{Y(1)}^{-1}(\theta)|X\right)-\theta\right)\right]}{f_{Y(0)}\left(F_{Y(0)}^{-1}(\theta)\right)f_{Y(1)}\left(F_{Y(1)}^{-1}(\theta)\right)},$$

$$\Omega_2^S=E\left[F_{Y(0)}\left(F_{Y(0)}^{-1}(\theta)|X\right)\left(1-F_{Y(0)}\left(F_{Y(0)}^{-1}(\theta)|X\right)\right)f_X(X)/f_X\left(X|T=0\right)\right],$$

$$\Omega_3^S=E\left[F_{Y(1)}\left(F_{Y(1)}^{-1}(\theta)|X\right)\left(1-F_{Y(1)}\left(F_{Y(1)}^{-1}(\theta)|X\right)\right)f_X(X)/f_X\left(X|T=1\right)\right].$$

COROLLARY: $\widehat{QTET}$ and $\widehat{QTE}$ achieve the efficiency bounds derived by Firpo (2005).

The proofs of Theorem 2 and its corollary can be found in the appendices B and C. Although the estimators in this section are based on nonparametric methods, they are $\sqrt{n}$ consistent because the first-step infinite dimensional estimates are integrated over all observations to obtain the finite-dimensional second step estimate. The average derivative quantile regression estimator of Chaudhuri, Doksum and Samarov (1997) is similar in this aspect. The asymptotic equivalence of the sample quantile and $\hat{q}_t^S(\theta)$ could be surprising but the reason is clear: asymptotically, the bandwidth is zero and no assumption is made about the dependence between *Y* and *X*. Note that if *Y* is linear in *X*, then the parametric estimator uses the optimal, infinite bandwidth while the nonparametric estimator constrains the bandwidth to go to zero.

[15] avar and acov are the normalized asymptotic variances and covariance given above.

This explains why the parametric estimator is more efficient than the nonparametric one in this case. The efficiency gain of the estimator of Section 3 results from the parametric assumptions. If the parametric restrictions are satisfied, we increase precision; if they are not satisfied, the estimator may be inconsistent.

When comparing the asymptotic variances of $\hat{q}_c$ and $\hat{q}_c^S$, we note that both consist of two parts and that both first parts are exactly identical. This is the variance that we would obtain if we knew the true conditional quantiles and, therefore, this part does not depend on the method used to estimate the conditional quantiles. The second part is the contribution of the first step estimation to the second step variance which differs between $\hat{q}_c$ and $\hat{q}_c^S$. While $X_i\hat{\beta}_0(\tau)$ and $X_j\hat{\beta}_0(\tau')$ are correlated, $X_i\hat{\beta}_0(\tau, X_i)$ and $X_j\hat{\beta}_0(\tau', X_j)$ are asymptotically independent if $X_i \neq X_j$ because the coefficients are only locally estimated. Thus, we do not need to account for these covariances and the double integral appearing in the asymptotic variance of $\hat{q}_c$ disappears for $\hat{q}_c^S$. Finally, we can use the form of the asymptotic variance of $\hat{\beta}_0(\tau, X_i)$ to simplify $\text{avar}\left(\hat{q}_c^S\right)$.

We show in the corollary of Theorem 2 that $\widehat{QTET}$ and $\widehat{QTE}$ achieve the semiparametric efficiency bounds without knowledge of the propensity score derived by Firpo (2005). Moreover, he proves that his propensity score weighting estimators also achieve the semiparametric efficiency bounds. Thus, both estimators of quantile treatment effects are asymptotically equivalent, just as the Heckman, Ichimura and Todd (1998) and the Hirano, Imbens and Ridder (2003) estimator of average treatment effects. Naturally, their finite sample properties may be very different. The relative advantages of both approaches are discussed in the conclusion.

The estimation of the asymptotic variance of $\widehat{QTET}$ and $\widehat{QTE}$ is in principle simpler to estimate than that of the parametric estimators. We only need to estimate unconditional distribution (and quantile) functions and unconditional densities. Unconditional distributions and quantile functions are estimated by (9) and (10) respectively. For the estimation of unconditional distributions, we use kernel density estimates with Silverman (1986) bandwidth. If we must estimate the density of an unobserved distribution, we use the principle described in (8) for the parametric estimator, that is we apply a kernel density estimator on the estimated unobserved distribution.

5. Monte-Carlo simulations

Asymptotic results are interesting partly because we hope that they describe approximately the behavior of the estimators in finite-samples. In this section we try to find out how the proposed estimators behave in finite samples. We first study the parametric estimator, which we call QQR (quantile based on quantile regression). Then we compare it to the estimator proposed by Machado and Mata (2005) since this estimator is frequently applied. Finally, we consider the estimator based on nonparametric first step quantile regression (QNQR). Software to implement the proposed estimators in *R* and to replicate the Monte Carlo simulation are available at the author's website.[16]

5.1. Parametric estimator

We consider a simple model with three correlated covariates and a constant:

$$Y(t) = 1 + X_1 + X_2 + X_3 + \varepsilon(t)(1 + X_1) \qquad t = 0,1$$

where $X_1 \sim U(0,1)$, $X_2 \sim B(0.5)$, $X_3 \sim N(0,1)$, $\text{cor}(X_1, X_2) = 0.4$, $\text{cor}(X_1, X_3) = 0.49$, $\text{cor}(X_2, X_3) = 0.4$, $\varepsilon(0) \sim t(1)$, $\varepsilon(1) \sim N(0,1)$ and $\Pr(T = 1|X) = 0.5$. The distribution of the covariates and the median coefficients do not depend on the treatment status but the error term is normally distributed for the treated and Cauchy distributed for the control units. Thus, the quantile treatment effect is positive below the median and negative above. It also allows us to compare the behavior of the estimators in the presence of a standard normal and an extremely fat tailed distribution.

We consider 3 different sample sizes $n_0 = n_1$: 100, 400 and 1600 and we set the number of replications to 10000, 5000 and 2500, respectively. We report the results for $\hat{q}_0(\theta)$, $\hat{q}_1(\theta)$ and $\widehat{QTET} = \hat{q}_1(\theta) - \hat{q}_c(\theta)$, both evaluated at 3 different quantiles: 5%, 25% and 50%. Table 1 reports the bias, standard error, skewness, kurtosis and mean squared error (MSE) of the estimates. The relative MSE of the sample quantile is also given for $\hat{q}_0(\theta)$ and $\hat{q}_1(\theta)$ in order to evaluate the efficiency gains achieved by the QQR.

As expected, the bias is smaller in the center of the distribution and with normal error terms. In the cases where there is a bias, it tends to disappear as the sample size increases. The analytically established convergence rate of the estimator is confirmed since quadrupling the

[16] *R* is an open-source programming environment for conducting statistical analysis and graphics that can be downloaded at no cost from the site www.r-project.org.

sample size cuts the standard errors by half and the MSE by about 75%. Considering the skewness and kurtosis, the distribution of the estimates appears to be already fairly close to the normal distribution with 100 observations for the median. A higher sample size is necessary for lower quantiles in the presence of Cauchy distributed error terms, but the convergence to the values of the normal distribution is clear. Finally, the QQR is almost always more efficient than the sample quantile. The only exception arises for the 5th percentile with small sample sizes and Cauchy distributed error terms.

We also evaluate the performance of the analytical estimator for the variance proposed in Section 3.3 and compare its performance with that of the bootstrap. In order to keep the computation time reasonable, the results for the bootstrap are based on only 4000, 2000 and 1000 replications for sample sizes of 100, 400 and 1600 respectively. Within each Monte Carlo replication, 100 bootstrap replications were drawn. We present results only for the *QTET* but they are representative for the results of other estimands. Table 2 gives different criterions that allow us to evaluate the estimators. It reports first the rejection frequencies by a Wald test of the true null hypothesis for 3 different confidence levels. Secondly, since this first evaluation does not allow us to evaluate the precision of the estimates, the median bias and the median absolute deviation from the true value for both estimators are also given. We take the empirical standard errors obtained in the Monte Carlo simulations as the "true" values.[17]

The empirical sizes of the tests confirm that both the analytic estimator and the bootstrap are consistent for the standard error of the QQR. With the exception of the 5th percentile with low sample sizes, both are reasonable estimators with empirical sizes near the theoretical ones. If we consider the MAD of both estimators, we note that the analytic estimator is more precise than the bootstrap (with 2 exceptions). Thus, the analytic estimator of the variance is not only faster to compute but also more efficient and its use in applications can be recommended.

5.2. Comparison to Machado and Mata (2005) estimator

Machado and Mata (2005, MM hereafter) also propose using quantile regression in order to estimate counterfactual unconditional wage distributions. Their estimator is widely used in various applications, see for instance Albrecht, Björklund and Vroman (2003), Melly (2005a) and Autor, Katz and Kearney (2005a and 2005b). However, no asymptotic results and no

[17] Another possibility would be to compute the asymptotic standard error analytically, but what we want is to estimate the empirical variance of the estimate and not the asymptotic variance.

method to estimate the variance consistently have been provided[18]. We show in this section that the MM estimator is numerically identical to our estimator if the number of simulations goes to infinity and, thus, the results of our paper apply also to their estimator.

The idea underlying their technique is the probability integral transformation theorem. If U is uniformly distributed on $[0,1]$, then $F^{-1}(U)$ has F as distribution function. Thus, for a given X_i and a random $\theta \sim U[0,1]$, $X_i\beta_0(\theta)$ has the same distribution as $Y(0)|X_i$. If we draw a random X from the control population instead of keeping X_i fixed, $X\beta_0(\theta)$ has the same distribution as $Y(0)|T=0$. Formally, the procedure proposed by MM involves 4 steps:

1. Generate a random sample of size m from a $U[0,1]$: $u_1,...,u_m$.
2. Estimate m different quantile regression coefficients: $\hat{\beta}_0(u_i)$, $i=1,...,m$.
3. Generate a random sample of size m with replacement from $\{X_i\}_{T_i=0}$, denoted by $\{\tilde{X}_i\}_{i=1}^{m}$.
4. $\{\tilde{Y}_i = \tilde{X}_i\hat{\beta}_0(u_i)\}_{i=1}^{m}$ is a random sample of size m from the unconditional distribution of $Y(0)|T=0$.

Naturally, alternative distributions could be estimated by drawing X from another distribution and using different coefficient vectors. As noted by Autor, Katz and Kearney (2005a), this procedure is equivalent to numerically integrating the estimated conditional quantile functions over the distributions of X and θ. The principles of the MM estimator and of the QQR are identical. First, since the observations are assumed to be *iid*, the QQR uses all observations instead of a single one with each of the m different quantile regression coefficients. Second, if $m \to \infty$, the probability that a coefficient $\hat{\beta}_0(\tau_j)$ is chosen is exactly equal to $\tau_j - \tau_{j-1}$ since for all $\tau_{j-1} \le u_i \le \tau_j$, $\hat{\beta}_0(u_i) = \hat{\beta}_0(\tau_j)$ and $\Pr(\tau_{j-1} \le u_i \le \tau_j) = \tau_j - \tau_{j-1}$. In other words, if $m \to \infty$, the MM estimator is numerically identical to the QQR.

A Monte-Carlo simulation illustrates this result. We keep the same data-generating process as in Section 5.1 and estimate $q_c(0.5)$ in 5000 replications using a sample of 400 observations[19]. Figure 1 plots the correlation between the MM estimator and the QQR as a function of m. The equality of both estimators when $m \to \infty$ is clear. Figure 2 shows that the imperfect

[18] Albrecht, Van Vuuren and Vroman (2004) derive the asymptotic distribution under the special assumption that the number of replications is of the same order as the number of observations. Therefore, they obtain different results. Their assumption entails the efficiency of the estimator, as explained below.

[19] Other quantiles, sample sizes or estimands lead exactly to the same conclusions.

correlation between the QQR and the MM estimator is simply due to the noise added by the bootstrap procedure of MM. The MSE of the QQR is always lower than the MSE of MM but both converge if $m \to \infty$. Almost all applications of the MM procedure set *m* equal to the sample size. We note that the MSE of the MM estimator for $m = n_1 = n_0 = 400$ is more than twice as large as the MSE of the QQR and, thus, the efficiency loss is really important in most of the applications.

Table 4 shows that the bias of the MM estimator does not depend on *m*, as expected, but the standard errors of the estimates diminishes as we increase the number of replications. Thus, a large number of replications is necessary in order to obtain good MSE properties. Naturally, estimating a large number of replications is time consuming especially when the number of observations is high and the estimation of the whole quantile regression process is not possible. QQR can be computed faster and uses the information contained in the data more efficiently. Simulation procedures are useful if there is no analytical solution to the problem. However, they are not necessary if we can, as in our case, use moment conditions in order to derive an analytical estimator for the parameters of interest.

5.3. Semiparametric estimator

We now present the results of a Monte-Carlo simulation using nonparametric quantile regression in the first-step. We consider a nonlinear model with a single regressor and a constant. The error term is again hit by a linear heteroscedastic scale. Formally

$$Y(t) = 5 + X + 4\cos(X) + 2\sin(3X) + \varepsilon(t)(0.5 + |X|) \qquad t = 0,1$$

where $X|T=0 \sim N(0,4)$, $X|T=1 \sim N(0,1)$, $\varepsilon(0) \sim t(1)$ and $\varepsilon(1) \sim N(0,1)$.

We consider 3 different quantiles: 5%, 25% and 50% and 4 different sample sizes $n_0 = n_1$: 100, 400, 1600 and 6400. The number of replications was set to 8000, 4000, 2000 and 1000, respectively. We use an Epanechnikov kernel and estimate 100 quantile regressions at each observation. Choosing a bandwidth for a semi-parametric estimator is a difficult task since the bandwidth does not appear in the first-order approximation of the asymptotic distribution. Here, it is even harder because we must choose not only one but a large number of bandwidths: one for each quantile regression. We make the simplifying assumptions of Yu and Jones (1998), which implies that the optimal bandwidth[20] for one quantile can be derived

[20] This is the optimal bandwidth for the nonparametric estimator and therefore cannot be the optimal bandwidth for the second step estimator. However, we can hope that this is a sensible bandwidth once we have corrected for the convergence rate. In any case, the asymptotic properties are still valid without the optimal bandwidth.

from the optimal bandwidth for another quantile and we are left with the choice of a single bandwidth. We set the bandwidth of the median regression quite arbitrarily to $n^{-1/4}\,\mathrm{sd}(X_i)$.

Table 4 reports the bias, the standard errors, the mean squared error (MSE), the skewness and the kurtosis of $\hat{q}_0^S$, $\hat{q}_1^S$ and $\widehat{QTET}$. The relative mean squared error of the sample quantile is also given for $\hat{q}_0^S$ and $\hat{q}_1^S$. The consistency, the convergence rate and the asymptotic normality of the estimates are confirmed by the Monte Carlo simulations but more observations are needed when the error terms are Cauchy distributed than when they are normally distributed. The relative MSE of the sample quantiles converges to 1 as predicted by the asymptotic results. Once again we note a difference between $\hat{q}_0^S$ and $\hat{q}_1^S$: in finite samples, the QNQR tends to have a higher MSE than the sample quantiles in the presence of Cauchy disturbances while it tends to have a smaller MSE in the presence of normal disturbances. Table 5 evaluates the analytic estimator of the variance by using the same criteria as in Table 2. It was not possible to bootstrap the results because of the computation time. Analytical standard errors tend to be close to the observed standard errors and fairly precise. With at least 400 observations the empirical sizes are close to the nominal ones. These results lead us to conclude that the proposed procedures constitute a complete system for estimating *QTE*s and for making consistent inference.

6. Applications: black-white wage differentials

As explained in the introduction and in Section 5.2., several estimators similar to the QQR have already be applied in different contexts: Gosling, Machin and Meghir (2000), Albrecht, Björklund and Vroman (2004), Machado and Mata (2005), Autor, Katz and Kearney (2005a and 2005b), for instance. In this section, we show in another application how the estimation of *QTE* complements the estimation of *ATE* and how the semiparametric estimator allows us to relax too restrictive assumptions.

Race differentials in labor market outcomes remain persistent. Although earnings appeared to converge during most of the postwar period, the black-white wage gap has now stagnated for the last two decades. We complement the traditional decomposition of the racial wage gap (see Altonji and Blank, 1999, for a survey) by considering the wage gap at different points of the distribution, which allows us to answer different questions about the racial wage gap. We can test several hypothesis like the presence of a glass ceiling or of sticky floors. Usually, the literature has identified the existence of a glass ceiling when the pay gap is significantly larger

at the top of the distribution. Arulampalam, Booth and Bryan (2005) identify a sticky floor when the wage gap is significantly larger at the bottom of the wage distribution. Both hypotheses have been put forward as explanations for the black/white wage gap. Some scholars have argued that blacks have become increasingly divided into two economic worlds: the emerging black middle class that rejoins the white middle class and the excluded black underclass, left out of the white economic world. This sticky floor hypothesis should appear in our results as an in absolute value decreasing black wage gap as we move along the wage distribution. Alternatively, if black employees are being discriminated against in promotion, that is if black employees have a lower probability of being promoted to jobs with higher responsibilities even if they have the same ability distribution as the white employees, then we should observe a glass ceiling pattern, i.e. a higher racial wage gap at the top of the distribution.

We use data from the Merged Outgoing Rotation Groups of the Current Population Survey for the year 2001. We restrict the sample to men who are between 16 and 65 years old. To simplify the analysis, we simply multiply the censored observations by 1.33. This has virtually no effect on the results since less than 1% of the observations are censored. An alternative would be to estimate censored quantile regression. We consider the differences in log wage between white and black workers and define $T_i = 0$ for white and $T_i = 1$ for black. Descriptive statistics for the variables of interest are given in Table 6. The covariates consist of education, potential experience and three regional dummies (*south* is the excluded category). The means of the relevant variables show that black workers are less educated, slightly more experienced and concentrated in the South region.

6.1. Parametric estimator

Figure 3 plots the decomposition (6) of the black wage gap with a 95% confidence interval obtained by the analytical estimator of Section 3.3. The estimated total differential shows that the black wage gap is higher at the high end of the distribution than at the lower end. This could be interpreted as an indicator for the glass ceiling phenomenon. However, this could also arise from different distributions of characteristics for white and black. In fact, after correcting for the effects of characteristics, we find that the black wage gap is first increasing but is then stable from the 30^{th} percentile until the end of the distribution. We cannot really interpret this pattern as a glass ceiling effect since we would expect the race gap to increase particularly at the high end of the distribution. Thus, none of the two hypothesis (glass ceiling

and sticky floor) is verified and we observe a lower racial wage gap at the low end of the distribution. We see two possible explanations for this pattern. Discrimination is probably more difficult to justify[21] for very basic jobs, where all employees are doing the same task. Customer discrimination is maybe also less relevant for some low-paid jobs, which are occupied predominantly by black workers.

This decomposition depends crucially on the parametric assumption for consistency. A simple test of the functional form can be performed by comparing the sample quantiles with the quantiles implied by the linear quantile regression model. Figure 4 plots the differences between both estimates for white[22] workers with a 95% bootstrap confidence interval. It is obvious that the model is misspecified with too high estimates in the extreme parts of the distribution and too low estimates in the middle of the distribution. For the majority of quantiles the differences are significantly different from 0. In order to suppress the parametric assumption, we now estimate the first step nonparametrically.

6.2. Semiparametric estimator

Since there are only 11 different values for education and 4 different regions, we can use exact nonparametric matching on these variables and must smooth only over experience. In this dimension, we use the same kernel and bandwidths as in Section 5.3. By looking at Figure 4, we can now check if the quantiles implied by the model and the raw quantiles are similar. The differences are now flat and not U-shaped any more as it was the case for the parametric first-step. Therefore, we trust these results more than those of Section 6.1.

The decomposition plotted in Figure 5 does not really contradict the above interpretation. The analytically estimated standard errors are higher and the estimates are less smooth but the main message remains unchanged. The different distribution of characteristics explains about one third of the level in wages and a large part of the glass ceiling pattern. Neither a glass ceiling nor a sticky floor phenomenon can be observed but the racial wage discrimination is lower at the lowest part of the distribution.

7. Conclusion

This paper proposes and implements parametric and semiparametric procedures to estimate unconditional distributions in the presence of covariates. This allows us to estimate

[21] In order to avoid a lawsuit.
[22] The differences are not significantly different from 0 for black workers, but the sample size is much lower.

counterfactual distributions and quantile treatment effects. The estimators are based on the estimation of the conditional distribution by parametric or nonparametric quantile regression. The first step estimates are then integrated over the range of the covariates in order to obtain the unconditional distribution. $\sqrt{n}$ consistency and asymptotic normality of both estimators are shown and analytical procedures to estimate their variances are provided. We also show that the parametric estimator of unconditional distributions is more precise than the sample quantile[23] and that the semiparametric estimator of quantile treatment effects achieves the efficiency bound. Monte-Carlo simulations show that the asymptotic results are useful approximations in medium sample sizes. We apply the proposed estimators to decompose the black-white gap in earnings and find no glass ceiling effect for blacks.

The estimators proposed in this paper are based on the unconfoundedness assumption. In order to estimate quantile or average treatment effects, three types of estimators have been proposed: the regression estimators, the matching (in the restrictive way) estimators and the estimators using the propensity score. Our estimators are clearly of the first type since we estimate the conditional distribution function by quantile regression. If fully nonparametric procedures are used, all approaches yield numerically identical results. However, in applications, a fully nonparametric approach is often not possible and the different restrictions will have different effects on the estimation. The more we go into the parametric direction, the more the choice of the approach matters. If the sample size is too small or if the number of covariates is too high, the two tractable competitors are the propensity score matching and the QQR. While propensity score matching estimators assume that $p(X)$ satisfies a parametric distributional assumption, the QQR assumes that we know up to a finite number of parameters how Y depends on X. It depends on the application in question which of these assumptions is more likely to be satisfied. We have a preference for the second type of assumptions because they are often easier to interpret[24] and because no distributional assumption is necessary[25].

New directions of research naturally arise from this paper. The efficiency of the parametric estimator can certainly be improved by using weighted quantile regression (Zhao 1999). It would be interesting to investigate if this weighted estimator attains an efficiency bound. A method to choose the bandwidths is the most urgent development needed to fully specify the

[23] Naturally, the sample quantiles can only be used to estimate observed distributions.
[24] The coefficients have a natural interpretation as rates of return to the human capital characteristics. Theoretical models can help to choose the parametric specification.
[25] Probit or logit estimators are consistent only if the latent error term is normally respectively logistically distributed.

estimator using nonparametric quantile regression. The optimal choice of smoothing parameters is a problem appearing in the implementation of a lot of semiparametric estimators proposed during the last decade. We must say that no fully satisfying solution has so far been developed. An additional problem, which is specific to the proposed estimator, is that the optimal bandwidth probably depends on the quantile of the regression and, thus, a huge number of different bandwidths must be chosen. The computational burden may simply be too high for a large range of methods, and simplifying assumptions, such as the ones used in Section 5.3, may be unavoidable.

Appendix A: Proof of theorem 1

Theorems 1 and 2 are applications of the results of Chen, Linton and Van Keilegom (2003, CLV hereafter). They extend the results of Newey (1994) and Andrews (1994) for non-smooth objective functions, allowing for a non-parametric first step estimation. We follow, as much as possible the notation of CLV but we must replace their θ by q since θ already symbolizes the quantile of interest in the quantile regression framework. We derive the asymptotic distribution of $\hat{q}_c(\theta)$. The other results can be derived similarly. Define

$$
\begin{aligned}
&Z_i = (Y_i, X_i, T_i),\\
&m(X_i, q, \beta(\cdot)) = \theta - \int_0^1 1(X_i'\beta(\tau) - q \le 0)\, d\tau,\\
&M(q, \beta(\cdot)) = \underset{T=1}{E}\left[m(X, q, \beta(\cdot)) \right],\\
&M_n(q, \beta(\cdot)) = n_1^{-1} \sum_{i:T_i=1} m(X_i, q, \beta(\cdot)).
\end{aligned}
$$

The moment condition *M* is satisfied since, at the true parameters $\beta_0(\cdot)$ and q_c,

$$
\begin{aligned}
&M(q_c, \beta_0(\cdot)) = \underset{T=1}{E}\left[m(X, q_c, \beta_0(\cdot)) \right] = \underset{T=1}{E}\left[\theta - \int_0^1 1(X\beta_0(\tau) - q_c \le 0)\, d\tau \right]\\
&= \theta - \underset{T=1}{E}\left[\Pr(Y(0) < q_c \,|\, X) \right] = \theta - \underset{T=1}{E}\left[F_{Y(0)}(q_c \,|\, X) \right] = \theta - F_{Y(0)}(q_c \,|\, T=1) = 0.
\end{aligned}
$$

The asymptotic distribution of the first step parametric quantile regression process has been derived by Gutenbrunner and Jureckova (1992):

$$
\sqrt{n}\left(\hat{\beta}_0(\tau) - \beta_0(\tau) \right) \Rightarrow b(\tau),
$$

where $b(\cdot)$ is a mean zero Gaussian process with covariance function:

$$
\operatorname{cov}\left[b(\tau) b(\tau') \right] = \left(\min(\tau, \tau') - \tau\tau' \right) D_1^t(\tau)^{-1} D_0^t D_1^t(\tau)^{-1}. \tag{12}
$$

The consistency of $\hat{q}_c$ is straightforward to show and is based on the consistency of the quantile regression coefficients. Thus, we concentrate on the asymptotic normality and examine the 6 conditions of theorem 2 in CLV.

Condition (2.1): $\hat{q}_c$ is, by definition, the θ^{th} quantile of the sample $\left\{ \left\{ X_i \hat{\beta}_0(\tau_j) \right\}_{j=1}^{J} \right\}_{i:T_i=1}$ where each "observation" is weighted by $\tau_j - \tau_{j-1}$. Koenker and Bassett (1978) show that quantiles can also be defined as solutions to optimization problems. $\hat{q}_c$ solves

$$\arg\min_q n_1^{-1}\sum_{i:T_i=1}\sum_{j=1}^{J}\left(\tau_j-\tau_{j-1}\right)\rho_\theta\left(X_i\hat{\beta}\left(\tau_j\right)\le q\right)=\arg\min_q n_1^{-1}\sum_{i:T_i=1}\int_0^1\rho_\theta\left(X_i\hat{\beta}\left(\tau\right)\le q\right)d\tau\,.$$

$M_n\left(q,\hat{\beta}\right)$ is the derivative of this problem. Koenker and Bassett (1978) show in theorem 3.3 that $M_n\left(\hat{q}_c,\hat{\beta}\right)=o\left(n^{-1/2}\right)$ and, thus, satisfies condition (2.1).

Condition (2.2): $\Gamma_1=\dfrac{\partial M\left(q_c,\beta_0(\cdot)\right)}{\partial q}=\dfrac{\partial}{\partial q}\underset{T=1}{E}\left[\theta-\int_0^1 1\left(X\beta_0\left(\tau\right)-q_c\le 0\right)d\tau\right]$

$$=-\frac{\partial}{\partial q}\underset{T=1}{E}\left[F_{Y(0)}\left(q_c\left|X\right.\right)\right]=-\frac{\partial F_{Y(0)}\left(q_c\left|T=1\right.\right)}{\partial q}=-f_{Y(0)}\left(q_c\left|T=1\right.\right).$$

By assumption *P.vii.* $f_{Y(0)}\left(q_c\left|T=1\right.\right)$ is continuous and not zero and thus conditions 2 (i) and (ii) of CLV are satisfied.

Condition (2.3): $\Gamma_2\left(q,\beta_0(\cdot)\right)\left[\hat{\beta}_0(\cdot)-\beta_0(\cdot)\right]$, the pathwise derivative of $M\left(q,\beta_0(\cdot)\right)$, exists and $\Gamma_2\left(q,\beta_0(\cdot)\right)\left[\hat{\beta}_0(\cdot)-\beta_0(\cdot)\right]=\underset{T=1}{E}\left[f_{Y(0)}\left(q\left|X\right.\right)X\left(\hat{\beta}_0\left(F_{Y(0)}\left(q\left|X\right.\right)\right)-\beta_0\left(F_{Y(0)}\left(q\left|X\right.\right)\right)\right)\right]$

since $\dfrac{\partial m\left(X_i,q,\beta_0(\cdot)\right)}{\partial\beta\left(\tau_k\right)}=\dfrac{\partial}{\partial\beta\left(\tau_k\right)}\left[\theta-\int_0^1 1\left(X_i\beta_0\left(\tau\right)-q\le 0\right)d\tau\right]=0$ if $\tau_k\ne F_{Y(0)}\left(q\left|X_i\right.\right)$ and

$\dfrac{\partial m\left(X_i,q,\beta_0(\cdot)\right)}{\partial\beta\left(\tau_k\right)}=-\dfrac{\partial}{\partial\beta\left(\tau_k\right)}\int_0^1 1\left(X_i\beta_0\left(\tau\right)-q\le 0\right)d\tau=f_{Y(0)}\left(q\left|X_i\right.\right)X_i$ if $\tau_k=F_{Y(0)}^{-1}\left(q\left|X_i\right.\right)$.

The last equality follows from $\int_0^1 1\left(X_i\beta_0\left(\tau\right)\le q\right)d\tau=F_{Y(0)}\left(q\left|X_i\right.\right)$. Now, by assumption *P.iv.*, $f_{Y(0)}\left(q_0\left|X\right.\right)$ is continuous and $f_{Y(0)}\left(q_0\left|X\right.\right)'$ is bounded. Moreover, by assumption *P.vii.*, X_i is bounded. Thus condition 3 (i) is satisfied with $c=\sup\left|f_{Y(0)}\left(q_c\left|X\right.\right)'X\right|$ and condition 3 (ii) is also satisfied for the same reasons.

Condition (2.4): The firs step estimator is a parametric, $\sqrt{n}$ consistent estimator and thus satisfies condition (2.4).

Condition (2.5): We verify conditions (2.5) by applying theorem 3 of CLV. By definition:

$$\left|m\left(X_i,q',\beta'(\cdot)\right)-m\left(X_i,q,\beta(\cdot)\right)\right|^2\le\left|\int_0^1\left(1\left(X_i'\beta'(\tau)\le q'\right)-1\left(X_i'\beta(\tau)\le q\right)\right)d\tau\right|$$

$$\le\left|\int_0^1\left(1\left(X_i'\beta'(\tau)\le q'\right)-1\left(X_i'\beta(\tau)\le q'\right)\right)d\tau\right|+\left|\int_0^1\left(1\left(X_i'\beta(\tau)\le q'\right)-1\left(X_i'\beta(\tau)\le q\right)\right)d\tau\right|.$$

We consider only the last term of the sum of the above right hand side, since the other term can be treated similarly (using the fact that X_i is bounded by assumption *P.vii.*):

$$\begin{aligned}
&\sup_{|q'-q|\le\delta} E\left\{\left|\int_0^1 \left(1\left(X\beta(\tau)\le q'\right)-1\left(X\beta(\tau)\le q\right)\right)d\tau\right|\right\} \\
&\le E\left\{\left|\int_0^1 \left(1\left(X\beta(\tau)\le q+\delta\right)-1\left(X\beta(\tau)\le q-\delta\right)\right)d\tau\right|\right\} \\
&\le E\left[F_{Y(0)}\left(q+\delta\middle|X\right)-F_{Y(0)}\left(q-\delta\middle|X\right)\right]\le K\delta
\end{aligned}$$

for some $K<\infty$, where the last inequality is due to the assumption that $\sup_{u,X}\left|f_{Y(0)}\left(u\middle|X\right)'\right|<\infty$. Hence condition 3.2 is satisfied with $r=2$ and $s=1/2$ and condition 3.3 holds by remark 3(ii) in CLK.

Condition (2.6): We now verify condition (2.6') which implies condition 2.6. Condition (2.6') (i) is trivially satisfied: $M_n\left(q_c,\beta_0(\cdot)\right)=n_1^{-1}\sum_{i:T_i=1} m\left(X_i,q_c,\beta_0(\cdot)\right)$, $\underset{T=1}{E}\left[m\left(X,q_c,\beta_0(\cdot)\right)\right]=0$ (shown above) and $\underset{T=1}{\text{var}}\left[m\left(X,q_c,\beta_0(\cdot)\right)\right]=\underset{T=1}{E}\left[\left(\theta-F_{Y(0)}\left(q_c\middle|X\right)\right)^2\right]\le 1$.

In order to verify condition (2.6') (ii) remember that:

$$\Gamma_2\left(q_c,\beta_0(\cdot)\right)\left[\hat\beta_0(\cdot)-\beta_0(\cdot)\right]=E\left[f_{Y(0)}\left(q_c\middle|X\right)X\left(\hat\beta_0\left(F_{Y(0)}\left(q_c\middle|X\right)\right)-\beta_0\left(F_{Y(0)}\left(q_c\middle|X\right)\right)\right)\right].$$

We now substitute in the Bahadur representation for $\hat\beta_0\left(F_{Y(0)}\left(q_c\middle|X\right)\right)-\beta_0\left(F_{Y(0)}\left(q_c\middle|X\right)\right)$, interchange integral and summation an approximate to obtain

$$\begin{aligned}
&\Gamma_2\left(q_c,\beta_0(\cdot)\right)\left[\hat\beta_0(\cdot)-\beta_0(\cdot)\right] \\
&=n_1^{-1}\sum_{i:T_i=1} f_{Y(0)}\left(q_c\middle|X_i\right)X_i\,\underset{T=0}{E}\left[f_{Y(0)}\left(F_{Y(0)}^{-1}\left(F_{Y(0)}\left(q_c\middle|X_i\right)\right)\middle|X\right)XX'\right]^{-1} n_0^{-1}\sum_{j:T_j=0}\omega_{ij}+o_p\left(n^{-0.5}\right) \\
&=n_1^{-1}\sum_{i:T_i=1}\psi(Z_i)+o_p\left(n^{-0.5}\right)
\end{aligned}$$

where $\omega_{ij}=X_j\left(F_{Y(0)}\left(q_c\middle|X_i\right)-1\left(Y_j\le X_j\beta_0\left(F_{Y(0)}\left(q_c\middle|X_i\right)\right)\right)\right)$ and $n_0^{-0.5}\sum_{j:T_j=0}\omega_{ij}$ is asymptotically normally distributed with mean 0 and variance $\underset{T=0}{E}\left[F_{Y(0)}\left(q_c\middle|X_i\right)\left(1-F_{Y(0)}\left(q_c\middle|X_i\right)\right)XX'\right]$. Therefore, $E\left(\psi(Z_i)\right)=0$ and $Var\left(\psi(Z_i)\right)<\infty$ by

assumptions *P.iv* and *P.vii*.

We can now derive $V_1 \equiv E\left[\left(m\left(X_i,q_c,\beta_0\right)+\psi\left(Z_i\right)\right)\left(m\left(X_i,q_c,\beta_0\right)+\psi\left(Z_i\right)\right)'\right]$. Note first that $m\left(X_i,q_c,\beta_0\right)$ and $\psi\left(Z_i\right)$ are uncorrelated. We have already derived $\text{var}\left[m\left(X_i,q_c,\beta_0\left(\cdot\right)\right)\right]=\underset{T=1}{E}\left[\left(\theta-F_{Y(0)}\left(q_c\left|X_i\right.\right)\right)^2\right]$. Then, using the notation introduced in (12), we find that $\text{var}\left(\psi\left(Z_i\right)\right)$ is equal to

$$\int\int f_{Y(0)}\left(q_c\left|x\right.\right)f_{Y(0)}\left(q_c\left|z\right.\right)x'\text{cov}\left[\hat{\beta}_0\left(F_{Y(0)}\left(q_c\left|x\right.\right)\right),\hat{\beta}_0\left(F_{Y(0)}\left(q_c\left|z\right.\right)\right)\right]zdF_X\left(x\left|T=1\right.\right)dF_X\left(z\left|T=1\right.\right)$$

or, integrating over the quantiles instead than over the distribution of *X*,

$$\int_0^1\int_0^1 E\left[f_{Y(0)}\left(q_c\left|X\right.\right)X'\left|F_{Y(0)}\left(q_c\left|X\right.\right)=\tau\right.\right]\text{cov}\left[\hat{\beta}(\tau),\hat{\beta}(\tau')\right]E\left[f_{Y(0)}\left(q_c\left|X\right.\right)X\left|F_{Y(0)}\left(q_c\left|X\right.\right)=\tau'\right.\right]d\tau d\tau'.$$

Since all conditions of theorem 2 of CLV are satisfied, we apply this theorem and obtain (7). All other results of our theorem 1 can be derived similarly.

Appendix B: Proof of theorem 2

As for the parametric estimator, we derive only the asymptotic distribution of the estimator of the counterfactual quantile $\hat{q}_c^S$. Since the structure of the proof is basically the same for the estimator using parametric first step estimation, we discuss only the differences. Define

$$m^N\left(X_i,q,\beta\left(\cdot,X_i\right)\right)=\theta-\int_0^1 1\left(X_i\beta\left(\tau,X_i\right)-q\le 0\right)d\tau.$$

Conditions (2.1), (2.2), (2.3) and (2.5) can be verified in the same way as for the parametric estimator, replacing $\beta(\tau)$ by $\beta\left(\tau,X_i\right)$. Assumptions *S.iv.*, *S.v.* and *S.vi.* ensure that the bias of the nonparametric quantile regression goes to zero faster than $n^{-1/2}$ and that the convergence rate of $\hat{\beta}_0\left(\tau,X_i\right)$ is at least of order $n^{-1/4}$, thus satisfying condition 2.4. In order to verify condition (2.6') (ii), we use the Bahadur representation for $X_i\left(\hat{\beta}_0\left(\tau,X_i\right)-\beta_0\left(\tau,X_i\right)\right)$ derived by Chaudhuri (1991):

$$\frac{1}{n_0h_n}\sum_{j:T_j=0}\frac{\tau-1\left(Y_j\le F_{Y(0)}^{-1}\left(\tau\left|X_i\right.\right)\right)}{f_X\left(X_i\left|T=0\right.\right)f_{Y(0)}\left(F_{Y(0)}^{-1}\left(\tau\left|X_i\right.\right)\right|X_i\right)}K\left(\frac{X_j-X_i}{h_n}\right)+o_p\left(n^{-0.5}\right).$$

We obtain then the following representation:

$$\Gamma_2\left(q_c,\beta_0\left(\cdot\right)\right)\left[\hat{\beta}_0\left(\cdot\right)-\beta_0\left(\cdot\right)\right]$$

$$= \underset{T=1}{E}\left\{ f_{Y(0)}\left(q_c|X\right)\left[\hat{\beta}_0\left(X, F_{Y(0)}\left(q_c|X\right)\right) - \beta_0\left(X, F_{Y(0)}\left(q_c|X\right)\right)\right]\right\}$$

$$= n_1^{-1} \sum_{i:T_i=1} \frac{f_{Y(0)}\left(q_c|X_i\right)}{n_0 h_n f_X\left(X_i|T=0\right)} \sum_{j:T_j=0} K\left(\frac{X_j - X_i}{h_n}\right) \frac{\omega_{ij}^N}{f_{Y(0)}\left(q_c|X_i\right)} + o_p\left(n^{-0.5}\right)$$

$$= n_1^{-1} \sum_{i:T_i=1} \frac{1}{n_0 h_n f_X\left(X_i|T=0\right)} \sum_{j:T_j=0} K\left(\frac{X_j - X_i}{h_n}\right) \omega_{ij}^S + o_p\left(n^{-0.5}\right)$$

$$= n_1^{-1} \sum_{i:T_i=1} \psi^S\left(Z_i\right) + o_p\left(n^{-0.5}\right)$$

where $\omega_{ij}^S = \left(F_{Y(0)}\left(q_c|X_i\right) - 1\left(Y_j \le X_j\beta_0\left(F_{Y(0)}\left(q_c|X_i\right)\right), X_i\right)\right)$. Given X_i, ω_{ij}^S is asymptotically normal with mean zero and variance $F_{Y|X}\left(q_0|X_i\right)\left(1 - F_{Y|X}\left(q_0|X_i\right)\right)$. Therefore, $E\left(\psi^S\left(Z_i\right)\right) = 0$ and

$$\text{var}\left(\psi^S\left(Z_i\right)\right) = \underset{T=1}{E}\left[\frac{f_X\left(X|T=1\right)}{f_X\left(X|T=0\right)} F_{Y(0)}\left(q_c|X\right)\left(1 - F_{Y(0)}\left(q_c|X\right)\right)\right],$$

which is finite by the overlap assumption. Thus, we can apply theorem 2 of CLV and we obtain (11). The other results from our theorem 2 can be proven similarly. For instance, using the same procedure as for $\hat{q}_c^S$, we obtain

$$\sqrt{n_1}\left(\hat{q}_1^S - q_1\right) \to N\left(0, \frac{\underset{T=1}{E}\left[\left(\theta - F_{Y(0)}\left(q_1|X\right)\right)^2\right] + \underset{T=1}{E}\left[F_{Y(1)}\left(q_1|X\right)\left(1 - F_{Y(1)}\left(q_1|X\right)\right)\right]}{f_{Y(1)}\left(q_1|T=1\right)^2}\right),$$

which is the variance of the θ^{th} quantile of $Y|T=1$ by the law of total variance.

Appendix C: Efficiency bounds

Firpo (2005) derives the efficiency bound for the *QTE* and the *QTET* assuming unconfoundedness, overlap and uniqueness of quantiles. Although his notation is almost totally different from our, we show in this appendix that the asymptotic variances of the proposed estimators $\widehat{QTET}$ and $\widehat{QTE}$ is equal to the efficiency bounds.

First, the efficiency bound for the *QTE* is given by Firpo (2005) on p. 9:

$$E\left[\frac{\text{var}\left[g_{1,\theta}\left(Y\right)|X,T=1\right]}{p\left(X\right)} + \frac{\text{var}\left[g_{0,\theta}\left(Y\right)|X,T=0\right]}{1-p\left(X\right)} + \left(E\left[g_{1,\theta}\left(Y\right)|X,T=1\right] - E\left[g_{0,\theta}\left(Y\right)|X,T=0\right]\right)^2\right]$$

where $g_{j,\theta} = -\left(\frac{1\left(Y \le q_{j,\theta}\right) - \theta}{f_{Y(j)}\left(q_{j,\theta}\right)} \right)$, $q_{j,\theta} = F^{-1}_{Y(j)}(\theta)$ for $j = 0,1$. Note that $\text{var}\left[g_{1,\theta}(Y) \middle| X, T=1 \right] = \frac{F_{Y(1)}\left(q_{1,\theta}|X\right)\left(1 - F_{Y(1)}\left(q_{1,\theta}|X\right)\right)}{f_{Y(1)}\left(q_{1,\theta}\right)^2}$ and $p(X) = \frac{f_X\left(X|T=1\right)}{f_X(X)} p$. Thus,

$$E\left[\frac{\text{var}\left[g_{1,\theta}(Y)|X,T=1 \right]}{p(X)} \right] = \int \frac{F_{Y(1)}\left(q_{1,\theta}|x\right)\left(1 - F_{Y(1)}\left(q_{1,\theta}|x\right)\right)}{p(x) f_{Y(1)}\left(q_{1,\theta}\right)^2} f_X(x)\,dx$$

$$= \int \frac{F_{Y(1)}\left(q_{1,\theta}|x\right)\left(1 - F_{Y(1)}\left(q_{1,\theta}|x\right)\right) f_X(x)}{f_{Y(1)}\left(q_{1,\theta}\right)^2 f_X\left(x|T=1\right) p} f_X(x)\,dx$$

$$= \frac{E\left[F_{Y(1)}\left(q_{1,\theta}|x\right)\left(1 - F_{Y(1)}\left(q_{1,\theta}|x\right)\right) f_X(x) / f_X\left(x|T=1\right) \right]}{f_{Y(1)}\left(q_{1,\theta}\right)^2 p}.$$

Similar calculations show that

$$\text{var}\left[g_{0,\theta}(Y)|X,T=0 \right] = \frac{E\left[F_{Y(0)}\left(q_{0,\theta}|X\right)\left(1 - F_{Y(0)}\left(q_{0,\theta}|X\right)\right) f_X(X) / f_X\left(X|T=0\right) \right]}{f_{Y(0)}\left(q_{0,\theta}\right)^2 p}.$$

Finally, with

$$E\left[\left(\underset{T=1}{E}\left[g_{1,\theta}(Y)|X \right] - \underset{T=0}{E}\left[g_{0,\theta}(Y)|X \right] \right)^2 \right]$$

$$= E\left[\underset{T=1}{E}\left[g_{1,\theta}(Y)|X \right]^2 + \underset{T=0}{E}\left[g_{0,\theta}(Y)|X \right]^2 - 2 \underset{T=1}{E}\left[g_{1,\theta}(Y)|X \right] \underset{T=0}{E}\left[g_{0,\theta}(Y)|X \right] \right]$$

$$= E\left[\frac{\left(F_{Y(1)}\left(q_{1,\theta}|X\right) - \theta \right)^2}{f_{Y(1)}\left(q_{1,\theta}|T=1\right)^2} + \frac{\left(F_{Y(0)}\left(q_{0,\theta}|X\right) - \theta \right)^2}{f_{Y(0)}\left(q_{0,\theta}|T=1\right)^2} - 2\frac{\left(F_{Y(1)}\left(q_{1,\theta}|X\right) - \theta \right)\left(F_{Y(0)}\left(q_{0,\theta}|X\right) - \theta \right)}{f_{Y(1)}\left(q_{1,\theta}\right) f_{Y(0)}\left(q_{0,\theta}\right)} \right]$$

$$= \frac{\text{var}\left[F_{Y(1)}\left(q_{1,\theta}|X\right) \right]}{f_{Y(1)}\left(q_{1,\theta}|T=1\right)^2} + \frac{\text{var}\left[F_{Y(0)}\left(q_{0,\theta}|X\right) \right]}{f_{Y(0)}\left(q_{0,\theta}|T=1\right)^2} - 2\frac{E\left[\left(F_{Y(1)}\left(q_{1,\theta}|X\right) - \theta \right)\left(F_{Y(0)}\left(q_{0,\theta}|X\right) - \theta \right) \right]}{f_{Y(1)}\left(q_{1,\theta}\right) f_{Y(0)}\left(q_{0,\theta}\right)}$$

we obtain the equality between the asymptotic variance of $\widehat{QTE}$ and the efficiency bound.

Now, for the $\widehat{QTET}$, Firpo derives the bound

$$E\left[\frac{p(X) V\left[g_{1,\theta|T=1}(Y)|X,T=1 \right]}{p^2} + \frac{p(X)^2 V\left[g_{0,\theta|T=1}(Y)|X,T=0 \right]}{p^2\left(1 - p(X)\right)} \right.$$

$$\left. + \frac{p(X)\left(E\left[g_{1,\theta|T=1}(Y)|X,T=1 \right] - E\left[g_{0,\theta|T=1}(Y)|X,T=0 \right] \right)^2}{p^2} \right].$$

Using the same results as for the *QTE* and noting that $p(X)=\frac{f_X(X|T=1)}{f_X(X)}p$, $1-p(X)=\frac{f_X(X|T=0)}{f_X(X)}(1-p)$ and $\frac{p(X)}{1-p(X)}=\frac{f_X(X|T=1)p}{f_X(X|T=0)(1-p)}$, we obtain

$$E\left[\frac{p(X)\operatorname{var}\left[g_{1,\theta|T=1}(Y)|X,T=1\right]}{p^2}\right]=\int\frac{p(x)F_{Y(1)}\left(q_{1,\theta|T=1}|x\right)\left(1-F_{Y(1)}\left(q_{1,\theta|T=1}|x\right)\right)}{p^2 f_{Y(1)}\left(q_{1,\theta|T=1}|T=1\right)^2}f_X(x)dx$$

$$=\int\frac{F_{Y(1)}\left(q_{1,\theta|T=1}|x\right)\left(1-F_{Y(1)}\left(q_{1,\theta|T=1}|x\right)\right)}{pf_{Y(1)}\left(q_{1,\theta|T=1}|T=1\right)^2}f_X(x|T=1)dx$$

$$=\underset{T=1}{E}\left[F_{Y(1)}(q_c|X)\left(1-F_{Y(1)}(q_c|X)\right)\right]\Big/f_{Y(1)}\left(q_{1,\theta|T=1}|T=1\right)^2$$

and

$$E\left[\frac{p(X)^2\operatorname{var}\left[g_{0,\theta|T=1}(Y)|X,T=0\right]}{p^2(1-p(X))}\right]=\int\frac{p(x)^2F_{Y(0)}\left(q_{0,\theta|T=1}|x\right)\left(1-F_{Y(0)}\left(q_{0,\theta|T=1}|x\right)\right)}{p^2(1-p(x))f_{Y(0)}\left(q_{0,\theta|T=1}|T=1\right)^2}f_X(x)dx$$

$$=\int\frac{F_{Y(0)}\left(q_{0,\theta|T=1}|x\right)\left(1-F_{Y(0)}\left(q_{0,\theta|T=1}|x\right)\right)f_X(x|T=1)}{f_{Y(0)}\left(q_{0,\theta|T=1}|T=1\right)^2f_X(x|T=0)(1-p)}f_X(x|T=1)dx$$

$$=\underset{T=1}{E}\left[F_{Y(0)}(q_c|X)\left(1-F_{Y(0)}(q_c|X)\right)f_X(X|T=1)/f_X(X|T=0)\right]\Big/f_{Y(0)}(q_c|T=1)^2.$$

Similar calculations show that

$$E\left[\frac{p(X)}{p^2}\left(E\left[g_{1,\theta|T=1}(Y)|X,T=1\right]-E\left[g_{0,\theta|T=1}(Y)|X,T=0\right]\right)^2\right]$$

$$=\frac{\underset{T=1}{\operatorname{var}}\left[F_{Y(0)}(q_c|X)\right]}{f_{Y(0)}(q_c|T=1)^2p}+\frac{\underset{T=1}{\operatorname{var}}\left[F_{Y(1)}(q_1|X)\right]}{f_{Y(1)}(q_1|T=1)^2p}-2\frac{\underset{T=1}{E}\left[\left(F_{Y(0)}(q_c|X)-\theta\right)\left(F_{Y(1)}(q_1|X)-\theta\right)\right]}{f_{Y(0)}(q_c|T=1)f_{Y(1)}(q_1|T=1)p}.$$

Thus, the asymptotic variance of $\widehat{QTET}$ is the same as the efficiency bound.

References

Abadie, A., Angrist, J. and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings", *Econometrica*, 70, 91-117.

Abadie, A., and G. I. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects", *Econometrica*, 74, 235-267.

Albrecht, J., A. Björklund, and S. Vroman (2003): "Is There a Glass Ceiling in Sweden?", *Journal of Labor Economics*, 21, 145-177.

Albrecht, J., A. Van Vuuren, and S. Vroman (2004): "Decomposing the Gender Wage Gap in the Netherlands with Sample Selection Adjustments", mimeo.

Altonji, J. G., and R. M. Blank (1999): "Gender and Race in the Labor Market", in *Handbook of Labor Economics*, Volume 3C, ed. by O. C. Ashenfelter and D. Card. New York: Elsevier Science Press, 3143-3259.

Andrews, D. W. K. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity", *Econometrica*, 62, 43-72.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006) "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure", *Econometrica*, forthcoming.

Arulampalam, W., A. L. Booth, and M. L. Bryan (2005): "Is there a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution", mimeo.

Autor, D. H., L. F. Katz, and M. S. Kearney (2005a): "Rising Wage Inequality: The Role of Composition and Prices", MIT mimeo.

Autor, D. H., L. F. Katz, and M. S. Kearney (2005b): "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists", MIT mimeo.

Blinder, A. (1973): "Wage Discrimination: Reduced Form and Structural Estimates", *Journal of Human Resources*, 8, 436-455.

Buchinsky, M. (1994): "Changes in US Wage Structure 1963-87: An Application of Quantile Regression", *Econometrica*, 62, 405-458.

Chaudhuri, P. (1991): "Nonparametric Estimates of Regression Quantiles and their Local Bahadur Representation", *Annals of Statistics*, 19, 760-777.

Chaudhuri, P., K. Doksum, and A. Samarov (1997): "On Average Derivative Quantile Regression", *Annals of Statistics*, 25, 715-744.

Chen, X., O. Linton, and I. Van Keilegom (2003): "Estimation of Semiparametric Models when the Criterion Function is not Smooth", *Econometrica*, 71, 1591-1608.

Chernozhukov, V. and C. Hansen (2006): "Instrumental quantile regression inference for structural and treatment effect models", *Journal of Econometrics*, forthcoming.

Chesher, A. (2003): "Identification in nonseparable Models", *Econometrica*, 71, 1405-1441.

Cowell, F. (2000): "Measurement of Inequality", in *Handbook of Income Distribution*, ed. by A. Atkinson, and F. Bourguignon. Amsterdam: North Holland, 87-166.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: a semiparametric approach", *Econometrica*, 65, 1001-1046.

Donald, S. G., D. A. Green, and H. J. Paarsch (2000): "Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates", *Review of Economic Studies*, 67, 609-633.

Firpo, S. (2005): "Efficient Semiparametric Estimation of Quantile Treatment Effects", mimeo.

Frölich M. (2005): "Propensity Score Matching without Conditional Independence Assumption", mimeo.

Gosling, A., S. Machin, and C. Meghir (2000): "The Changing Distribution of Male Wages in the UK", *Review of Economic Studies*, 67, 635-686.

Gutenbrunner, C., and J. Jurecková (1992): "Regression Rank-Scores and Regression Quantiles", *Annals of Statistics*, 20, 305-330.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects", *Econometrica*, 66, 315-331.

Hall, P., and S. Sheather (1988): "On the Distribution of a Studentized Quantile", *Journal of the Royal Statistical Society, Series B*, 50, 381-391.

Hausman, J. A. (1978): "Specification Test in Econometrics", *Econometrica*, 46, 1251-1271.

Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator", *The Review of Economic Studies*, 65, 261-294.

Heckman, J. J., J. Smith and N. Clements (1997): "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts", *Review of Economic Studies*, 64, 487-535.

Hendricks, W., and R. Koenker (1991): "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity", *Journal of the American Statistical Association*, 87, 58-68.

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

Imbens, G. I. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: a Review", *Review of Economic and Statistics*, 86, 4-29.

Juhn, C., K. M. Murphy, and B. Pierce (1993): "Wage Inequality and the Rise in Returns to Skill", *Journal of Political Economy"*, 101, 410-442.

Koenker, R. (2005): *Quantile Regression*. Cambridge: Cambridge University Press.

Koenker, R., and G. Bassett (1978): "Regression Quantiles", *Econometrica*, 46, 33-50.

Koenker, R., and S. Portnoy (1987): "L-Estimation for Linear Models", *Journal of the American Statistical Association*, 82, 851-857.

Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process", *Econometrica*, 70,1583-1612.

Lechner, M. (2001): "A Note on the Common Support Problem in Applied Evalutation Studies", University of St. Gallen, discussion paper 2001-01.

Lemieux, T. (2006): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skills", *American Economic Review*, forthcoming.

Machado, J. A. F., and J. Mata (2005): "Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression", *Journal of Applied Econometrics*, 20, 445-465.

Melly, B. (2005a): "Public-Private Sector Wage Differentials in Germany: Evidence from Quantile Regression", *Empirical Economics*, 30, 505-520.

Melly, B. (2005b): "Decomposition of Differences in Distribution using Quantile Regression", *Labour Economics*, 12, 577-590.

Melly, B. (2006): "Public and Private Sector Wage Distributions controlling for Endogenous Sector Choice", mimeo.

Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators", *Econometrica*, 62, 1349-1382.

Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9", translated in *Statistical Science*, 5 (1990), 465-480.

Oaxaca, R. (1973): "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, 14, 693-709.

Portnoy, S. (1991): "Asymptotic Behavior of the Number of Regression Quantile Breakpoints", *SIAM Journal of Scientific and Statistical Computing*, 12, 867-883.

Powell, J. L. (1984): "Least absolute deviations estimation for the censored regression model", *Journal of Econometrics*, 25, 303-325.

Powell, J. L., (1986): "Censored Regression Quantiles", *Journal of Econometrics*, 32, 143-155.

Siddiqui, M. (1960): "Distribution of Quantiles from a Bivariate Population", *Journal of Research of the National Bureau of Standards*, 64, 145-150.

Silverman, B.W. (1986): "*Density estimation*", London: Chapman and Hall.

Yu, K., and M. C. Jones (1998): "Local Linear Quantile Regression", *Journal of the American Statistical Association*, 93, 228-237.

Zhao, Q. (1999): "Asymptotically Efficient Median Regression in the Presence of Heteroscedasticity of Unknown Form", *Econometric Theory*, 17, 765-784.

Table 1: Monte Carlo simulation, parametric first step, point estimates.

Sample size	Bias	St. dev.	MSE	Skew.	Kurt.	Relative MSE of the sample quantile
			5^{th} percentile, control units, true value: -7.5032			
100	-2.0064	6.049	40.6123	-2.8884	21.2894	0.5156
400	-0.4751	2.1886	5.0148	-0.967	5.2957	0.888
1600	-0.1059	1.0357	1.0834	-0.498	3.4822	1.0279
			25^{th} percentile, control units, true value: -0.0795			
100	-0.0413	0.411	0.1706	-0.3001	3.3652	1.107
400	-0.0099	0.1986	0.0395	-0.1669	3.2084	1.1776
1600	-0.0038	0.1004	0.0101	-0.128	2.9951	1.154
			50^{th} percentile, control units, true value: 1.8887			
100.	0.0090	0.3273	0.1072	0.0661	3.0657	1.2014
400	0.0007	0.1640	0.0269	0.0085	2.9647	1.2251
1600	0.0003	0.0815	0.0066	-0.0403	2.9281	1.2555
			5^{th} percentile, treated units, true value: -1.2798			
100.	0.0193	0.3068	0.0945	-0.0128	2.9902	1.4539
400	0.0022	0.1528	0.0234	-0.0155	3.0263	1.4868
1600	0.0009	0.0754	0.0057	-0.0509	2.9009	1.5398
			25^{th} percentile, treated units, true value: 0.5125			
100	0.0131	0.2301	0.0531	0.0227	2.8731	1.3348
400	0.0002	0.1159	0.0134	-0.0162	2.9842	1.3453
1600	-0.0002	0.0564	0.0032	0.0191	3.0702	1.3089
			50^{th} percentile, treated units, true value: 1.8918			
100	0.0098	0.2320	0.0539	0.0041	2.9503	1.3240
400	0.0001	0.1152	0.0133	-0.0103	2.9435	1.3615
1600	0.0004	0.0574	0.0033	0.0071	3.1482	1.3525
			5^{th} percentile, *QTET*, true value: 6.2250			
100.	2.1150	6.2848	43.9678	3.0180	21.9114	
400	0.4812	2.2014	5.0768	0.9617	5.2372	
1600	0.1059	1.0374	1.0869	0.4756	3.4395	
			25^{th} percentile, *QTET*, true value: 0.5919			
100	0.0549	0.4377	0.1946	0.3465	3.4224	
400	0.0108	0.2068	0.0429	0.1663	3.1594	
1600	0.0036	0.1037	0.0108	0.0654	2.8879	
			50^{th} percentile, *QTET*, true value: 0.0031			
100	-0.0009	0.3443	0.1186	-0.0391	3.1260	
400	-0.0002	0.1674	0.0280	-0.0179	3.0266	
1600	-0.0001	0.0823	0.0068	0.0349	2.9636	

The number of replications is 2500, 5000 and 10000 respectively for 1600, 400 and 100 observations.

Table 2: Monte-Carlo simulation, parametric first step, estimation of the standard errors.

Estimator	Empirical size for a confidence level of 1%	5%	10%	Median bias	MAD
	5st percentile, 100 observations, "true" value: 6.2848				
analytic	0.0778	0.1251	0.1632	-2.1858	3.5338
bootstrap	0.039	0.0695	0.0935	-0.4586	3.6826
	5st percentile, 400 observations, "true" value: 2.2014				
analytic	0.0404	0.0810	0.1236	-0.2115	0.6732
bootstrap	0.0195	0.0450	0.0820	0.1052	0.5969
	5st percentile, 1600 observations, "true" value: 1.0374				
analytic	0.0240	0.0644	0.1148	-0.0475	0.1695
bootstrap	0.0210	0.0510	0.1030	-0.0071	0.1559
	25st percentile, 100 observations, "true" value: 0.4377				
analytic	0.0098	0.0413	0.0833	-0.0044	0.0571
bootstrap	0.0073	0.0323	0.0650	0.0372	0.0691
	25st percentile, 400 observations, "true" value: 0.2068				
analytic	0.0114	0.0530	0.0986	-0.0018	0.0131
bootstrap	0.0115	0.0535	0.1070	0.0026	0.0165
	25st percentile, 1600 observations, "true" value: 0.1037				
analytic	0.0100	0.0548	0.1032	-0.0025	0.0038
bootstrap	0.0090	0.0650	0.1030	-0.0019	0.0192
	50st percentile, 100 observations, "true" value: 0.3443				
analytic	0.0083	0.0438	0.0853	0.0072	0.0278
bootstrap	0.0060	0.0380	0.0848	0.0222	0.0370
	50st percentile, 400 observations, "true" value: 0.1674				
analytic	0.0114	0.0502	0.1000	-0.0018	0.0064
bootstrap	0.0100	0.0515	0.0990	0.0002	0.0104
	50st percentile, 1600 observations, "true" value: 0.0823				
analytic	0.0088	0.0472	0.1016	0.0001	0.0016
bootstrap	0.0120	0.0520	0.1070	-0.0001	0.0235

For the analytic estimator, the number of replications is 2500, 5000 and 10000 respectively for 1600, 400 and 100 observations. For the bootstrap estimator, the number of replications is 1000, 2000 and 4000 respectively for 1600, 400 and 100 observations.

Table 3: Monte-Carlo simulation, point estimates of $q_c(0.5)$ with 400 observations.

Estimator	Bias	Standard error	MSE	Relative MSE	Correlation
	true value: 1.8887				
MM with:					
$m = 100$	0.0074	0.3968	0.1575	5.8711	0.3855
$m = 400$	0.0014	0.2416	0.0584	2.1762	0.6689
$m = 1000$	-0.0006	0.2015	0.0406	1.5138	0.8167
$m = 10000$	0.0008	0.1683	0.0283	1.0552	0.9760
$m = 100000$	0.0007	0.1643	0.0270	1.0059	0.9976
$\hat{q}_c(0.5)$	0.0004	0.1638	0.0268	1.0000	1.0000

Results based on 5000 replications.

Table 4: Monte Carlo simulation, nonparametric first step, point estimates.

Sample size	Bias	St. dev.	MSE	Skew.	Kurt.	Relative MSE of the sample quantile
	5^{th} percentile, control units, true value: -9.9967					
100	-4.0656	11.1005	139.735	-3.993	33.843	0.2791
400	-2.4228	3.9373	21.3688	-1.2106	5.6593	0.3807
1600	-1.3499	1.6497	4.5423	-0.4672	3.3023	0.4303
6400	-0.5805	0.7238	0.8603	-0.2607	3.0978	0.5537
	25^{th} percentile, control units, true value: 2.4906					
100	-0.3502	1.2111	1.5892	-0.4817	3.2603	0.794
400	-0.1555	0.5905	0.3728	-0.3264	3.1843	0.9056
1600	-0.0678	0.2967	0.0926	-0.174	3.1047	0.9555
6400	-0.0169	0.1489	0.0224	-0.0896	2.9079	1.0227
	50^{th} percentile, control units, true value: 6.2308					
100	0.031	0.5568	0.3109	0.0107	3.2204	0.9467
400	0.0366	0.2691	0.0737	0.0688	2.9178	0.9985
1600	0.0205	0.1296	0.0172	-0.1041	3.0533	1.008
6400	0.0144	0.0646	0.0044	0.0381	2.9033	1.0051
	5^{th} percentile, treated units, true value: 3.7015					
100	0.0153	0.936	0.8763	-0.8658	4.296	0.9438
400	0.0042	0.443	0.1962	-0.447	3.4442	1.0759
1600	-0.0032	0.2247	0.0505	-0.1422	2.9996	1.0655
6400	0.0056	0.1126	0.0127	-0.0372	3.2399	1.0746
	25^{th} percentile, treated units, true value: 6.6709					
100	0.0555	0.3272	0.1101	-0.1129	3.0459	1.067
400	0.0316	0.1607	0.0268	0.0481	3.2409	1.0713
1600	0.014	0.0806	0.0067	-0.0026	2.9702	1.0525
6400	0.0088	0.0396	0.0016	-0.0718	2.9824	1.0211
	50^{th} percentile, treated units, true value: 8.5198					
100	0.0195	0.3479	0.1214	0.0825	3.1471	1.2787
400	0.0163	0.1833	0.0339	0.0612	3.0014	1.1576
1600	0.0074	0.092	0.0085	0.0528	2.9641	1.1052
6400	0.0037	0.0483	0.0023	0.0553	2.8465	1.0846
	5^{th} percentile, QTET, true value: 6.3287					
100	5.0822	27.7069	793.4025	30.2677	1372.56	
400	1.485	2.7951	10.0162	1.4287	6.5665	
1600	0.5487	1.0287	1.3588	0.4466	3.2192	
6400	0.2156	0.4775	0.2742	0.2293	3.0544	

Table 4 (cont.): Monte Carlo simulation, nonparametric first step, point estimates.

Sample size	Bias	St. dev.	MSE	Skew.	Kurt.	Relative MSE of the sample quantile
	25^{th} percentile, QTET, true value: 1.3986					
100	0.2217	0.7256	0.5756	1.0586	7.8828	
400	0.0819	0.276	0.0828	0.1927	3.1587	
1600	0.0293	0.1362	0.0194	0.1029	3.0828	
6400	0.0142	0.0688	0.0049	0.0764	3.0489	
	50^{th} percentile, QTET, true value: 0.8783					
100	-0.0225	0.5303	0.2817	-0.031	3.2076	
400	-0.0031	0.2459	0.0605	-0.0399	2.9529	
1600	0.0011	0.1232	0.0152	-0.0979	2.8415	
6400	0.004	0.0613	0.0038	-0.2134	2.8404	

The number of replications is 1000, 2000, 4000 and 8000 respectively for 6400, 1600, 400 and 100 observations.

Table 5: Monte Carlo simulation, nonparametric first step, estimation of the standard errors.

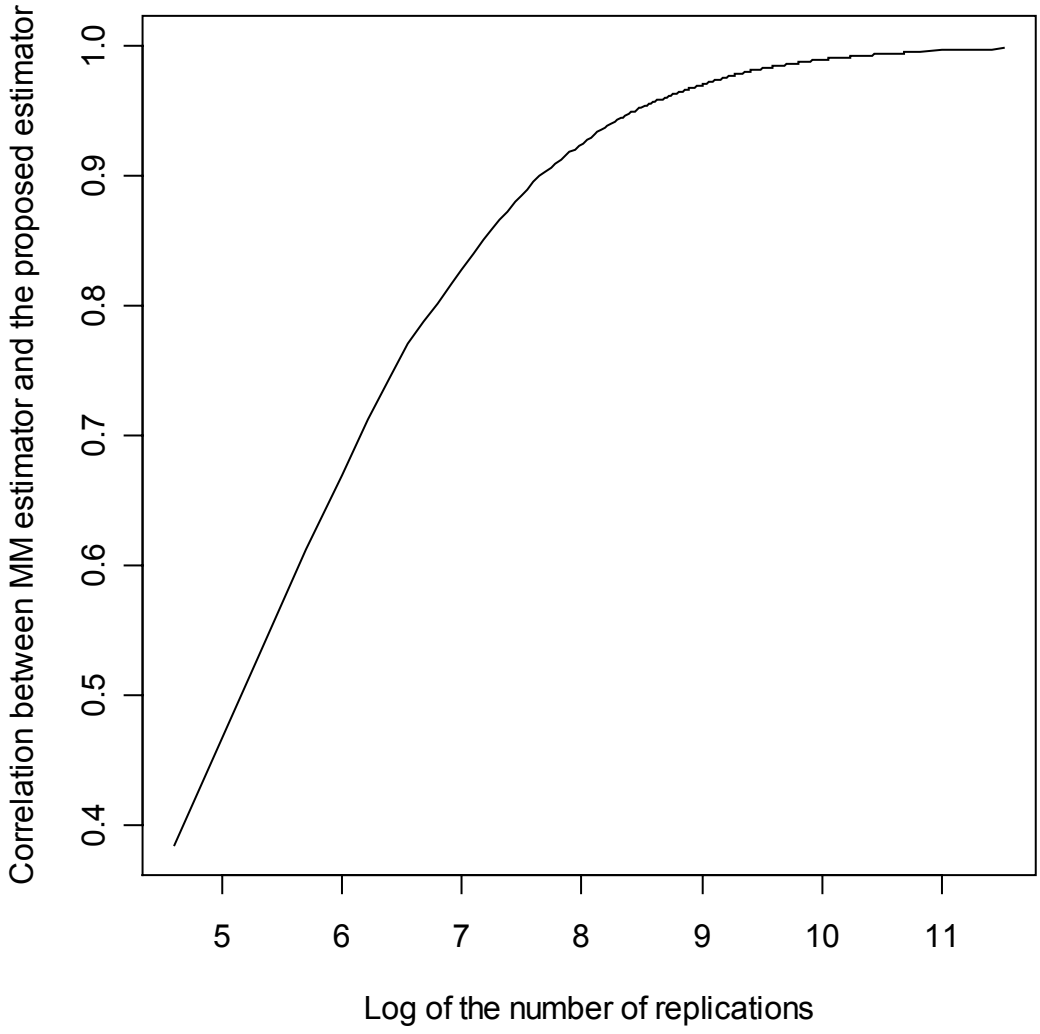
Sample size	Empirical size for a confidence level of 1%	5%	10%	"True" value	Median bias	MAE
			5st percentile			
100	0.1438	0.2300	0.2903	122.6105	-119.5533	119.5667
400	0.0230	0.0635	0.1065	2.7951	-0.6860	1.0394
1600	0.0090	0.0405	0.0860	1.0287	-0.0160	0.1694
6400	0.0100	0.0520	0.1070	0.4775	0.0059	0.0431
			25st percentile			
100	0.0078	0.0295	0.0534	0.7256	0.2345	0.2459
400	0.0048	0.0343	0.0768	0.2760	0.0204	0.0269
1600	0.0075	0.0400	0.0885	0.1362	0.0069	0.0087
6400	0.0080	0.0370	0.0990	0.0688	0.0016	0.0022
			50st percentile			
100	0.0076	0.0399	0.0790	0.5303	0.0159	0.0666
400	0.0043	0.0315	0.0675	0.2459	0.0220	0.0235
1600	0.0060	0.0330	0.0760	0.1232	0.0095	0.0095
6400	0.0070	0.0300	0.0700	0.0613	0.0048	0.0048

The number of replications is 1000, 2000, 4000 and 8000 respectively for 6400, 1600, 400 and 100 observations.

Table 6: Descriptive statistics, means

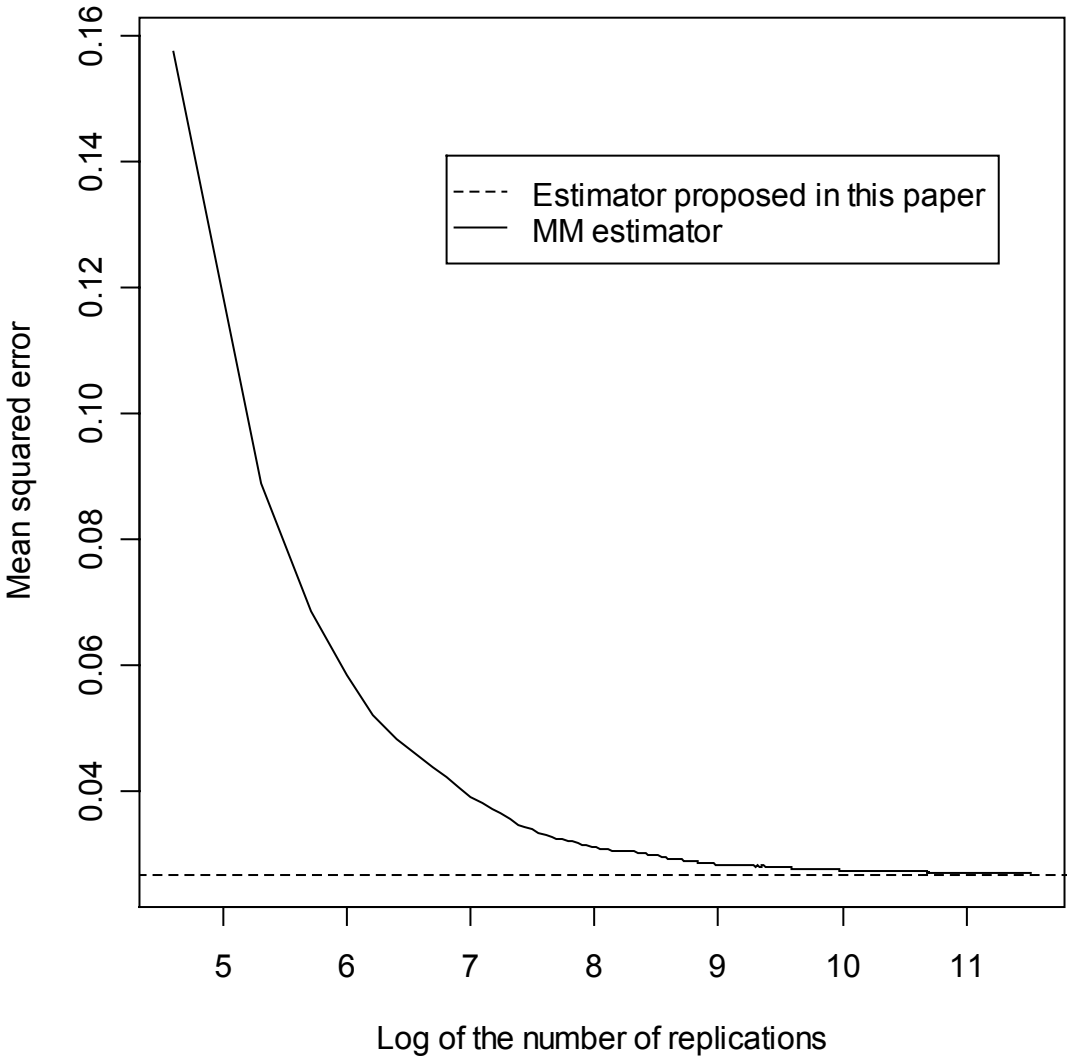
Variable	All	Whites	Blacks
Ln(wage)	2.1980	2.2159	1.9895
Experience	19.9547	19.9307	20.2333
Education	13.5631	13.6138	12.9750
South	0.2806	0.2560	0.5665
Midwest	0.2877	0.2972	0.1771
West	0.2325	0.2429	0.1115
Northeast	0.1993	0.2039	0.1449
Number of observations	40349	37147	3202

Figure 1: Correlation between the proposed estimator and the MM estimator as function of *m*.



Results based on 5000 Monte Carlo replications.

Figure 2: MSE of the proposed estimator and the MM estimator as a function of m.



Results based on 5000 Monte Carlo replications.

Figure 3: Decomposition of the black/white wage gap using parametric quantile regression.

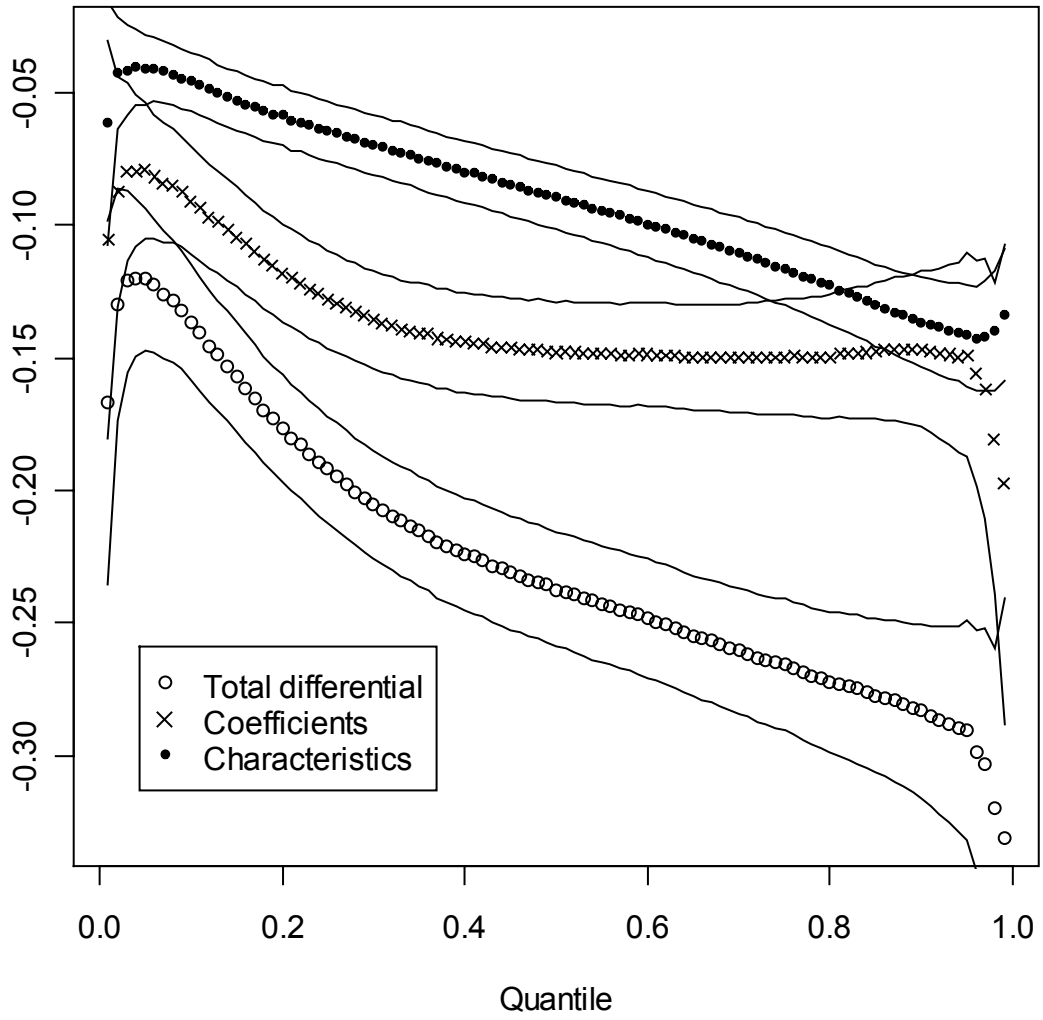
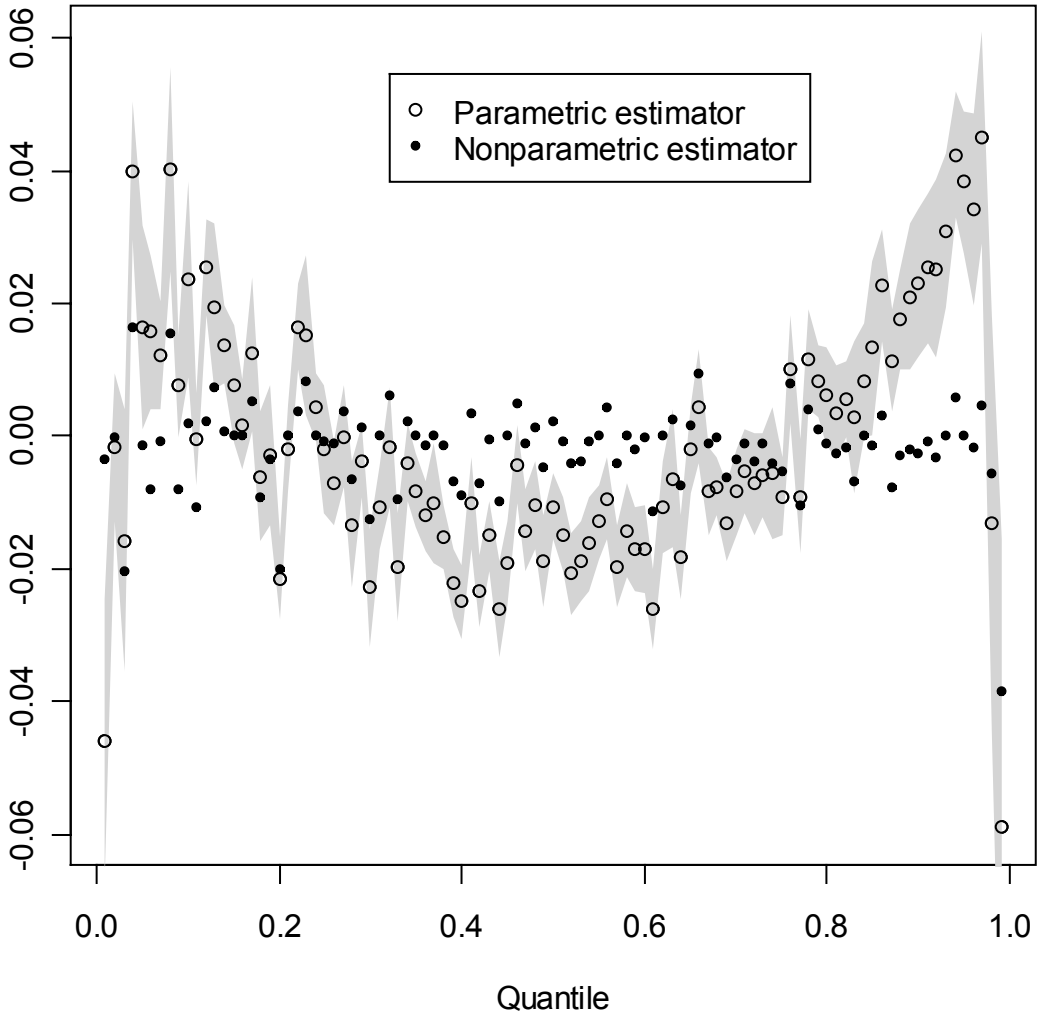


Figure 4: Difference between the unconditional quantiles implied by the model and the sample quantiles.



A 95% confidence interval obtained by bootstrapping the results 100 times is plotted.

Figure 5: Decomposition of the black wage gap using nonparametric quantile regression