

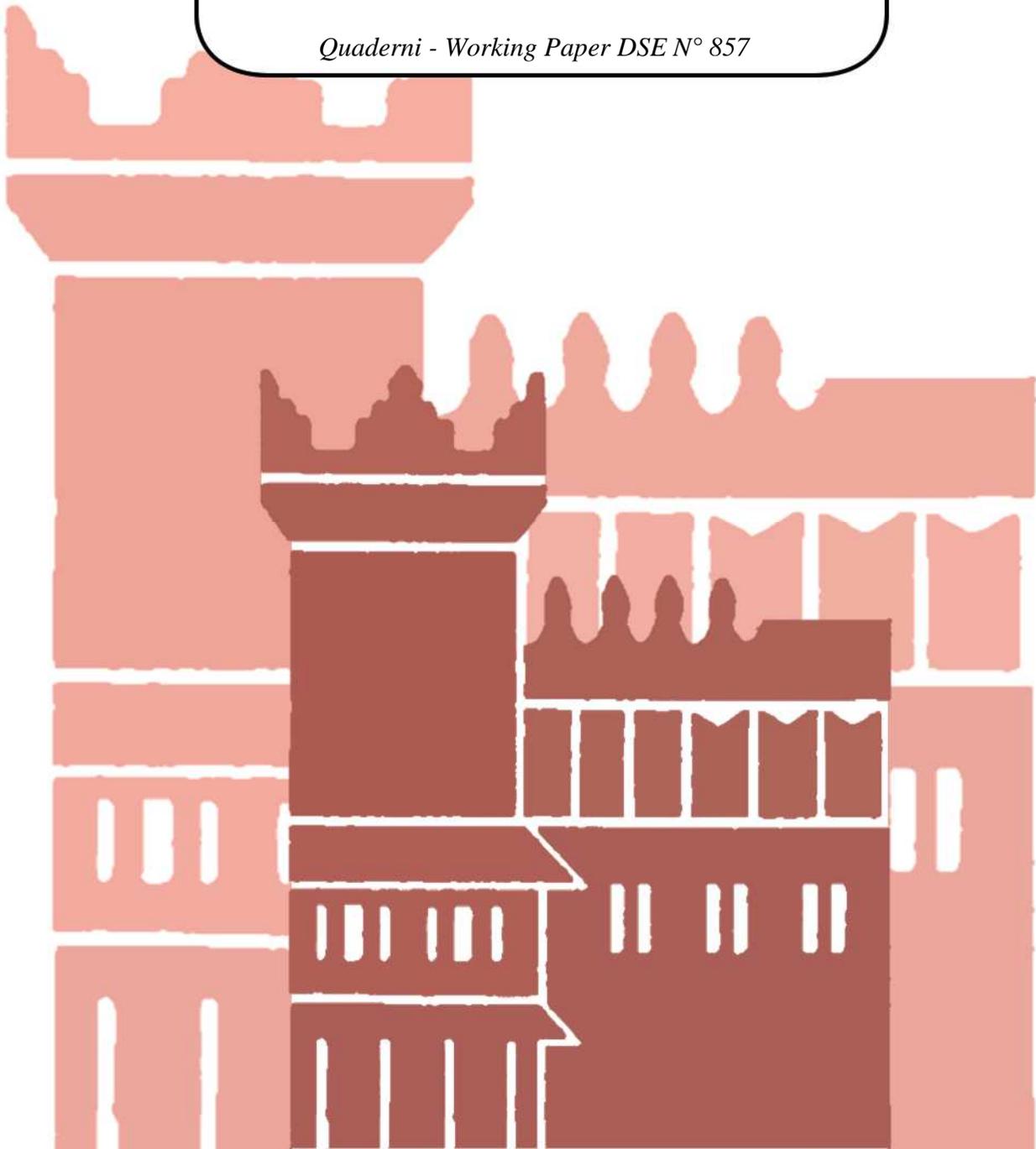


Alma Mater Studiorum - Università di Bologna  
DEPARTMENT OF ECONOMICS

**Unconditional and Conditional  
Quantile Treatment Effect:  
Identification Strategies and  
Interpretations**

Margherita Fort

*Quaderni - Working Paper DSE N° 857*



# Unconditional and Conditional Quantile Treatment Effect: Identification Strategies and Interpretations

Margherita Fort, University of Bologna)

This version: May 2012

## Abstract

This paper reviews strategies that allow one to identify the effects of policy interventions on the unconditional or conditional distribution of the outcome of interest. This distinction is irrelevant when one focuses on average treatment effects since identifying assumptions typically do not affect the parameter's interpretation. Conversely, finding the appropriate answer to a research question on the effects over the distribution requires particular attention in the choice of the identification strategy. Indeed, quantiles of the conditional and unconditional distribution of a random variable carry a different meaning even if identification of both these set of parameters may require conditioning on observed covariates.

**Keywords:** impact heterogeneity, quantile treatment effects, rank invariance.

JEL codes: C18

## 1 Introduction

In the recent years there has been a growing interest in the evaluation literature for models that allow essential heterogeneity in the treatment parameters and more generally for models that are informative on the impact distribution. The recent increase in the attention devoted to the identification and estimation of quantile treatment effects (QTEs) is due to their intrinsic ability to characterize the heterogenous impact of the treatment on various points of the outcome

---

\*Department of Economics, University of Bologna, IZA and CHILD; Piazza Scaravilli 2 Bologna, margherita.fort@unibo.it. Paper prepared for the 46th Italian Statistical Society meeting (invited). This paper benefited from comments by E. Rettore, B. Pacini and F. Mealli and participants to the 46th Italian Statistical Society meeting. Financial support of MIUR- FIRB 2008 project RBF089QQC-003-J31J10000060001 grant is gratefully acknowledged. The usual disclaimer applies.

distribution. QTEs are informative about the impact distribution when the potential outcomes observed under various levels of the treatment are comonotonic random variables. The variable describing the relative position of an individual in the outcome distribution thus plays a special role in this setting, representing at the same time the main dimension along which treatment effects are allowed to vary as well as a key ingredient to relate potential outcomes. Several identification approaches currently used in the literature for the assessment of mean effects have thus been extended to quantiles. Most of these strategies require to condition on a set of variables to achieve identification. While conditioning on a set of observed regressors does not affect the interpretation of the parameters in a mean regression, this is not the case for quantiles. The law of iterated expectations guarantees that the parameters of a mean regression have both a conditional and an unconditional mean interpretation. This does not carry over to quantiles, where conditioning on covariates affects the interpretation of the residual disturbance term. Indeed, since quantile regression allows one to characterize the heterogeneity of the treatment response only along this latter dimension, conditioning on covariates in quantile regression generally affects the interpretation of the results.

This paper reviews strategies aimed at identifying quantile treatment effects, covering strategies that deal with the identification of conditional and unconditional quantile treatment effects with particular attention to cross-sectional data applications in which the treatment is endogenous without conditioning on additional covariates. The aim of the paper is to provide useful guidance for users of quantile regression methods in choosing the most appropriate approach while addressing a specific research question.

The remainder of the paper is organized as follows. After introducing the basic notation and the key parameters of interest in Section 2, Section 3 reviews solutions to the identification of quantile treatment effects. The review covers strategies that are appropriate only when the outcome of interest is a continuous variable, i.e. in cases where the quantiles of the outcome distribution are unambiguously defined. It concludes illustrating some of the methods through two illustrative examples aimed at assessing the distributional impacts of training on earnings and of education on wages. Section 4 concludes.

## 2 What Are We After: Notation and Parameters of Interest

In this section I first introduce the notation used throughout the paper and then define the objects whose identification is sought.

$Y$  denotes the observed outcome,  $D$  the intensity of the treatment received and  $W$  a set of observable individual characteristics.  $W$  may include exogenous variables  $X$  and instruments  $Z$ .<sup>1</sup>  $Y$  is restricted to be continuous while  $D, W$  can be either continuous or discrete random variables. Both  $Y$  and  $D$  can be decomposed in two components: one of which is deterministic and one of which is stochastic. These two components need not be additively separable. The stochastic components account for differences in the distribution of  $D$  and  $Y$  across otherwise identical individuals. The econometric models reviewed in Section 3 place restriction on : i) the scale of  $D$ ; ii) the number of independent sources of stochastic variation in the model; iii) the distribution (joint, marginal, conditional) of these stochastic components and  $D$  or  $W \equiv (X, Z)$ ; iv) the scale of  $Z$ .  $Y_i^d$  denotes the *potential outcome* for individual  $i$  if the value of the treatment is  $d$ : it represents the outcome that would be observed had the individual  $i$  been exposed to level  $d$  of the treatment.  $F_{Y^d}(\cdot)$ ,  $f_{y^d}(\cdot)$  and  $F_{Y^d}^{-1}(\cdot) = q(d, \cdot)$  denote the corresponding cumulative distribution and density function and the quantile function. The conditional distribution and conditional quantile are denoted by  $F_{Y^d}(\cdot|x)$  and  $F_{Y^d}^{-1}(\cdot|x) = q(d, x, \cdot)$ .

We are interested in characterizing the dependence structure between  $Y$  and  $D$  eventually conditioning on a set of covariates  $W$  in the presence of essential heterogeneity and in the absence of general equilibrium effects. Knowledge of the joint distribution  $(Y^d)_{d \in \mathcal{D}}$  or the conditional joint distribution  $(Y^{d|x})_{d \in \mathcal{D}}$  would allow to characterize a distribution for the outcome for any possible level of the treatment. When potential outcomes are comonotonic, they can be described as different functions of the same (single) random variable and quantile treatment effects (QTEs) are informative on the impact distribution. The potential outcome could be written as  $y^d = q(d, u)$ ,  $u \sim \mathcal{U}(0, 1)$ ,  $q(d, u)$  is increasing in  $u$  as is refereed in the literature as the *structural quantile function*. If the potential outcomes are not comonotonic, QTEs are informative on the distance between potential outcomes distributions, which may be interesting *per se*, but not on the impact distribution. We thus concentrate on strategies that focus on

---

<sup>1</sup>Capital letters denote random variables and lower case letters denote realizations.

QTEs.<sup>2</sup> In the binary case, QTEs (see equation (2)) are defined as the horizontal distance between the distribution function in the presence and in the absence of the treatment ([9]; [15]).<sup>3</sup>

$$\delta(\tau) = F_{Y^1}^{-1}(\tau) - F_{Y^0}^{-1}(\tau) \quad 0 < \tau < 1 \quad (2)$$

We can distinguish conditional and unconditional quantile treatment effects by characterizing the uniformly distributed random variable that describes the quantile of the outcome variable. This distinction becomes clearer if we think about a specific empirical example.

### Motivating Example: Returns to Education or Training

There is a large literature that studies the returns to education. Key questions in this literature (e.g. does additional education cause wage increase? does additional schooling increase wages more for the more able than for the less able? does additional schooling increase or decrease wage inequality?) can be addressed using quantile regression methods. In this applications, the treatment is likely endogenous in the outcome equation without conditioning on additional covariates: typically researchers seek instruments that allow to isolate exogenous variation in education in the wage equation. Suppose we could measure the individual ability  $a_i$  that drives the endogeneity of education in the wage equation. Now, consider the alternative specifications for the wage model presented in equation (3), (4) where  $D$  denotes schooling (the treatment).

$$Y_i = \alpha_0(f(\varepsilon_i, a_i)) + \alpha_1(f(\varepsilon_i, a_i))D \quad (3)$$

$$Y_i = \beta_0(\varepsilon_i)a_i + \beta_1(\varepsilon_i)D \quad (4)$$

These specifications differ because they impose different structures of the variables governing the heterogeneity in the returns to education. In equation (3) the relative position of an individual in the wage distribution is determined by  $(\varepsilon_i, a_i)$ , i.e. by both an *unobserved* uniformly distributed error component  $\varepsilon_i$  and by the *observed* individual ability level while in equation (4) the relative position of the individual is only determined by  $\varepsilon_i$ . In both cases, we can think

---

<sup>2</sup>The review will not cover strategies that focus on other objects and may deliver QTEs as byproduct such as [8], for instance.

<sup>3</sup>In the continuous case  $\delta(\tau)$  represents the change in  $Y$  induced by a change in  $D$  from  $d$  to  $d + \epsilon$  when  $\epsilon$  is small.

$$\delta(\tau) = \frac{\partial Q_Y(\tau|d)}{\partial d} \quad 0 < \tau < 1 \quad (1)$$

Table 1: Moment conditions under assumptions in [2] and [11]

Quantile	conditional	unconditional
$Y_1$	$E[\{1(Y < q(1, x)) - \tau\} \cdot w_{y,d,x} \cdot D X] = 0$	$E[\{1(Y < q(1)) - \tau\} \cdot w_{y,d} \cdot D] = 0$
$Y_0$	$E[\{1(Y < q(0, x)) - \tau\} \cdot w_{y,d,x} \cdot (1 - D) X] = 0$	$E[\{1(Y < q(0)) - \tau\} \cdot w_{y,d} \cdot (1 - D)] = 0$
weight	$1 - \frac{D[1-P(Z=1 Y,D,X)]}{1-P(Z=1 X)} - \frac{(1-D)P(Z=1 Y,D,X)}{P(Z=1 X)}$	$E[\frac{Z-P(Z=1 X)}{P(Z=1 X)[1-P(Z=1 X)]} Y, D](2D - 1)$

Note: Positive weights are reported. See [2] and [11] for other definitions of weights.

about the relative position of an individual in the wage distribution as his/her *prone-ness* ([9]) to earn a high wage for a given level of schooling  $D$ . However, in model (3) we would refer to the *total prone-ness/ability* while in model (4) we would be speaking only about *unobserved prone-ness/ability*.<sup>4</sup> Using model (3) we can explore whether the returns to education vary depending on the individuals' *total* ability levels while using model (4) we can study how the returns to education vary for given observed ability levels. Individuals who earn high wages conditional on some specific level of ability may not be the same individuals who earn high wages in the sample. However conditioning on observed ability maybe important to be able to isolate the causal effect of schooling  $D$  on the distribution of wages  $Y$ . Equation (5) and Equation (6) represent the structural quantile function corresponding to model (3) and (4) respectively<sup>5</sup>: equation (5) is an example of an *unconditional* quantile regression model while equation (6) is an example of a *conditional* quantile regression model. This distinction might be empirically relevant since, in general, for a given  $\tau \in (0, 1)$ ,  $\alpha_1(\tau) \neq \beta_1(\tau)$ .

$$f(\varepsilon, a) \equiv \varepsilon^*, \varepsilon^* \sim \mathcal{U}(0, 1) \quad Q_Y(\tau|d) = \alpha_0(\tau) + \alpha_1(\tau)d \quad (5)$$

$$\varepsilon \sim \mathcal{U}(0, 1) \quad Q_Y(\tau|d) = \beta_0(\tau)a_i + \beta_1(\tau)d \quad (6)$$

### 3 Identification Strategies and Estimation

In cross-sectional applications, two main identification approaches have been extended to QTEs: strategies based on the *unconfoundedness assumption* and strategies based on the availability of an instrumental variable. In the first case, the researcher must be willing to assume that the joint distribution of the potential outcomes is independent of the treatment conditional on a set of exogenous

<sup>4</sup>To the best of my knowledge, [17] is the first to distinguish between *total* and *observed prone-ness*.

<sup>5</sup>Under comonotonicity of potential outcomes, the structural quantile function describes the link between potential outcomes.

covariates. Under this assumptions, conditional QTEs can be estimated as originally proposed by [14] and unconditional QTEs can be estimated as proposed by [10]. [2] and [6], [7] propose identifying assumptions for conditional quantiles when an instrumental variable is available. The assumptions of [2] guarantee identification of conditional and unconditional QTEs when the treatment is binary and endogenous and a binary instrument is available. They lead to the moment conditions described in Table 1: in both cases, identification relies on previous results ([1], [13]) that guarantee that in the subpopulation of *compliers* comparisons by treatment  $D$ , conditional on  $X$ , have a causal interpretation. Recall that *compliers* are individuals whose treatment status is affected by the instrument  $Z$  but that this sub-population cannot be identified directly from the data, because it is defined by means of potential outcomes. The moment conditions highlight that is possible to construct weights that 'find *compliers* in the population in an average sense' ([1]). The weights will differ when one is interested in the conditional or in the unconditional quantiles. Only the weights considered in the second case 'simultaneously balance the distribution of the covariates between treated and non-treated *compliers*' ([12]). In both cases weights are functions of  $P(Z = 1|X)$  and observed variables. Estimation thus proceeds in two steps: 1) weights are estimated; 2) weighted quantile regressions are run.<sup>6</sup> Estimation requires two steps also under the identification strategy proposed by [6], [7] and [17], [18] but does not involve re-weighting. The crucial assumption for identification in the approach by [6] is *rank invariance* or *rank similarity*, i.e. we require that the individual's rank in the potential outcome distribution, conditional on exogenous covariates, is not systematically affected by the treatment. The assumptions by [6] lead to the moment condition in equation (7). Equation (7) suggests an estimation procedure that first requires to compute the conditional quantiles of the random variable  $Y - q(d, x, \tau)$  given  $X$  and  $Z$ ; then, choose as estimate of  $q(d, x, \tau)$  the one that minimizes the absolute value of the coefficient associated with  $Z$  in the first step.<sup>7</sup>

$$Pr[Y - q(d, x, \tau) \leq 0|X, Z] = \tau. \quad (7)$$

---

<sup>6</sup>When identification is achieved relying on *uncounfoundedness*, the moment conditions are similar but the weights are identically 1 for conditional quantiles ([14]) and are  $\frac{D}{P(D=1|X)} + \frac{1-D}{1-P(D=1|X)}$  for unconditional quantiles ([10]).

<sup>7</sup>This approach can be used when the treatment and instrument are binary, discrete as well as continuous.

The instrumental variable approach for the identification of unconditional QTEs proposed by [17] delivers the moment conditions in equation (8)

$$E[Z\{1(Y \leq q(d, \tau)) - \tau_X\}] = 0, \tau_X \equiv P[Y \leq q(d, \tau)|X]. \quad E[1(Y \leq q(d, \tau)) - \tau] = 0. \quad (8)$$

These moment conditions reflect the idea that, first, the instrument  $Z$  does not affect the distribution of the disturbance once  $X$  is controlled for and, second, the joint distribution of  $X$  and the disturbance is unrestricted. Estimation involves first an estimation of the quantiles of  $Y - q(d, \tau)$  given  $X$  and  $Z$  and  $\tau_X$ ; then, a second step choose as estimate of  $q(d, \tau)$  the value that minimizes the coefficient of  $Z$  averaging over all possible values of  $X$ .

We now apply these strategies to two illustrative examples taken from the literature. Table 2 reports estimates of the effect of training (or education) on the conditional and unconditional distribution of earnings (or log wages) using data of males from [2] and data from [5], respectively.<sup>8</sup> Column (1) and (2) reports results delivered when training or education are treated as exogenous in the estimation of conditional and unconditional quantiles respectively. Column (3) and (4) report estimates that address the endogeneity of training or education in the outcome equation relying on [2]. These estimates apply to the sub-population of *compliers*. Column (5)-(8) report estimates based on [6] or [17]. These approaches guarantee global identification of conditional and unconditional QTEs. We discuss the top-panel estimates first: in the example from [2] the treatment assignment is randomized thus covariates are not needed for identification. Indeed, under both the identification approaches considered, training effects on the conditional and unconditional quantiles do not exhibit substantial differences in magnitude and all suggest that the effect of training is larger at the top of the earnings distribution.<sup>9</sup> In addition, both the identification strategies deliver similar results, suggesting that key assumptions are unlikely to be violated in both cases. Let's now turn to the estimates in the bottom part of

---

<sup>8</sup>In the second example, only reforms that increased compulsory schooling for 3 years are considered (i.e. only Greece, Italy and Finland) and the original treatment (years of education) and instrument (years of compulsory schooling) were recoded to binary. Estimates of column (1), (2), (3), (4) have been computed by the author using the STATA package `ivqte` by [12], except column (3) for the first example (taken from the article). Estimates in column (1) replicate original results in the papers except that standard errors are now robust to heteroskedasticity; estimates of columns (5)-(8) are taken from [18] for the AAI02 example and obtained using the STATA package `ivqreg` by Do Wan Kwack available from Christian Hansen's research page.

<sup>9</sup>When endogeneity of training is addressed, point estimates of the returns to training are generally lower in the unconditional distribution with respect to the returns observed holding race, age, education and marital status fixed.

Table 2: Effect of Training on the Conditional and Unconditional Distribution of Earnings ([2], males only) and Effect of Education on the Conditional and Unconditional Distribution of Log Wages ([5], males, Italy Greece and Finland, treatment and instrument recoded to binary)

Strategy	Exogenous Training		Endogenous Training					
	Conditional (1) KB78	Unconditional (2) F07	Monotonicity		Rank Invariance		Unconditional	
q			Conditional (3) AAI02	Unconditional (4) FM10	Conditional (5) CH08	(6) CH08 w/o controls	(7) P11 logit	(8) P11 probit
<b>Effect of Training on Earnings, Abadie et al. (2002) Obs. 5102</b>								
0.25	2510 417***	3058 (377)***	702 (670)	414 (754)	530 (629)	200 (746)	100 (753)	100 (750)
0.50	4420 (613)***	4678 (771)***	1544 (1074)	1291 (1239)	310 (1101)	1320 (1234)	790 (1151)	790 (1161)
0.75	4678 (901)***	4626 (1056)***	3131 (1376)**	2457 (1650)	2660 (1845)	1710 (1712)	1490 (1542)	1490 (1530)
0.85	4806 (1045)***	5532 (1241)***	3378 (1811)*	3971 (1886)**	3190 (1185)**	3580 (1427)**	3410 (1542)*	3410 (1550)*
<b>Effect of Education on Log Wages, Brunello et al. (2009) Obs. 2292</b>								
0.30	0.168 (0.024)***	0.223 (0.064)***	0.303 (0.142)**	0.514 (16.48)	0.836 (0.063)***	-0.198 (0.033)***	-	-
0.50	0.177 (0.024)***	0.208 (0.062)***	0.328 (0.126)***	0.521 (5.95)	0.985 (0.063)***	-5.119 (0.124)***	-	-
0.75	0.213 (0.026)***	0.297 (0.072)***	0.154 (0.168)***	0.599 (10.13)	1.868 (0.998)**	0.996 (0.037)***	-	-

Legend: Column labels refer to the estimation method. KB78: as in [14]; F07: as in [10]; AAI02: as in [2]; FM10: as in [11], [12]; CH08: as in [7]; P11: as in [18].

the table. In this example, covariates are needed for identification: we need to control for country specific secular trends in education and differences across countries in the levels of education and wages to be able to isolate the exogenous variation in education induced by school reforms. In this example, addressing endogeneity seems to have relevant consequences: the estimates in column (1) and (2) suggest that returns are increasing over the wage distribution, while estimates in columns (3) suggest the opposite -although precision of these estimates is low- and in column (4) we find no evidence of heterogeneity.<sup>10</sup> Estimates of conditional QTEs under rank invariance are reported in column (5); estimates of unconditional QTEs in column (6) assume rank invariance and do not use covariates for identification. The estimates in column (5) are unrealistic and suggest that rank invariance is unlikely to hold. Estimates in column (6) are negative and confirm that controlling for covariates is necessary for identification.

<sup>10</sup>In this example, we look at the effect of three more additional years of schooling on wages. Assuming linearity and dividing point estimates reported by three, the results in columns (1)-(3) are fairly consistent with the literature: association is lower than causal effects; causal estimate suggest a return between 10% and 4% for each additional year of education.

## 4 Conclusions

In this paper, I reviewed approaches that guarantee the identification of quantile treatment effects (QTEs). In many cases, these approaches correspond to extensions of strategies conventionally used in linear regression models (selection on observables, instrumental variables, fixed effects) to quantile regressions. An important consequence of the difference between the statistical tools applied in these two settings is that the interpretation of treatment parameters differs between conditional and unconditional quantile regressions, while, conversely, the law of iterated expectations guarantees that the treatment parameter in a linear regression as both a conditional and an unconditional mean interpretation. It is crucial to bear this in mind while using QTEs to answer a specific research question. Consider the recent proposal of [3] to link educator compensation to the ranks of their students within what the author call an *appropriately defined comparison sets*. The authors' suggest to employ methods in [4] to contrast actual ranks of students of a given teacher with some predicted counterfactual rank. Betebenner ([4]) however employs *conditional* quantile regression methods aimed specifically at answering questions like *Are there students with unusually low growth who need special attention?*, i.e. a value-added specification of the achievement. Barlevy and Neal ([3]) instead look for a method that allows to isolate the teachers contribution to a student rank in the achievement distribution in a given period, eventually conditioning on covariates for identification. In other words Barlevy and Neal would like to avoid attributing to a teacher the changes in performance of a student that are only due to his initial proficiency level. Standard value added specifications for students' achievement in quantile regression context are not the appropriate instrument to address questions about the heterogeneity in students' achievement depending on their initial ability level. Those quantile regression describe instead how students experiencing the largest gains in performance over a given time period perform relative to students experiencing the lowest gains in the same period. Cross-sectionally, some of the *high-gain* students may be in the lower part of the test score distribution.<sup>11</sup>

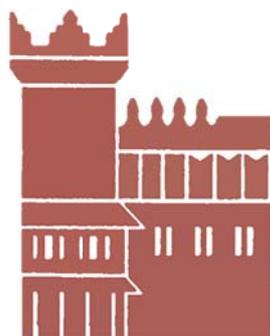
---

<sup>11</sup>A similar point was made by [16] in his discussion of the analysis of the effect of vouchers on student achievements.

## References

- [1] Abadie, A. : Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics* 113, 231-263 (2003)
- [2] Abadie, A. Angrist, J. and Imbens, G.: Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70, 91-117 (2002)
- [3] Barlevy, G. and Neal, D. : Pay for Percentile, *NBER Working Paper* 17194 (2010)
- [4] Betebenner, D. W.: Norm and Criterion-Referenced Student Growth. *Educational Measurement: Issues and Practice*, 28(4) 42-51 (2009)
- [5] Brunello, G. and Fort, M. and Weber, G.: Changes in Compulsory Schooling, Education and the Distribution of Wages in Europe. *Economic Journal* vol. 119(536), 516-539 (2009)
- [6] Chernozhukov, V. and Hansen, C. : An IV model of quantile treatment effects. *Econometrica* 73, 245-261 (2005)
- [7] Chernozhukov, V. and Hansen, C. : Instrumental variable quantile regression: A Robust Inference Approach. *Journal of Econometrics* 142(1), 379-398 (2008)
- [8] Chesher, A. : Identification in nonseparable models. *Econometrica* 71, 1405-1441 (2003)
- [9] Doksum, K.: Empirical probability plots and statistical inference for non-linear models in the two-sample case. *The Annals of Statistics* 2, 267-277 (1974)
- [10] Firpo, S.: Efficient semiparametric estimation of quantile treatment effect. *Econometrica* 75, 259-276 (2007)
- [11] Froelich, M. and Melly, B. (2010a) Unconditional Quantile Treatment Effects Under Endogeneity. IZA DP 3288 (2010)
- [12] Froelich, M. and Melly, B. : Estimation of Quantile Treatment Effects with STATA. *The Stata Journal* 10(3), 423-457 (2010b)
- [13] Imbens, G. and Rubin, D.: Estimating the Outcome Distribution for Compliers in Instrumental Variables Models. *Review of Economic Studies* 64, 555-574 (1997)

- [14] Koenker, R. and Bassett, G.S.: Regression Quantiles. *Econometrica* 46, 33-50 (1978)
- [15] Lehmann, E.H. : Nonparametrics: Statistical Models Based on Ranks. San Francisco,CA (1974)
- [16] Powell, D.: Unconditional Quantile Regression for Panel Data with Exogenous or Endogenous Treatment Variables. RAND Working Paper No. WR-710 (2010a)
- [17] Powell, D.: Unconditional Quantile Treatment Effects in the Presence of Covariates. RAND Working Paper No. WR-816 (2010b)
- [18] Powell, D.: Unconditional Quantile Regression for Exogenous or Endogenous Treatment Variables. RAND Working Paper No. WR-824 (2011)



Alma Mater Studiorum - Università di Bologna  
DEPARTMENT OF ECONOMICS

Strada Maggiore 45  
40125 Bologna - Italy  
Tel. +39 051 2092604  
Fax +39 051 2092664  
<http://www.dse.unibo.it>