Chapter 8. Quantile Regression and Quantile Treatment Effects

Joan Llull

Quantitative & Statistical Methods II
Master in Economics of Public Policy
Barcelona School of Economics

I. Introduction

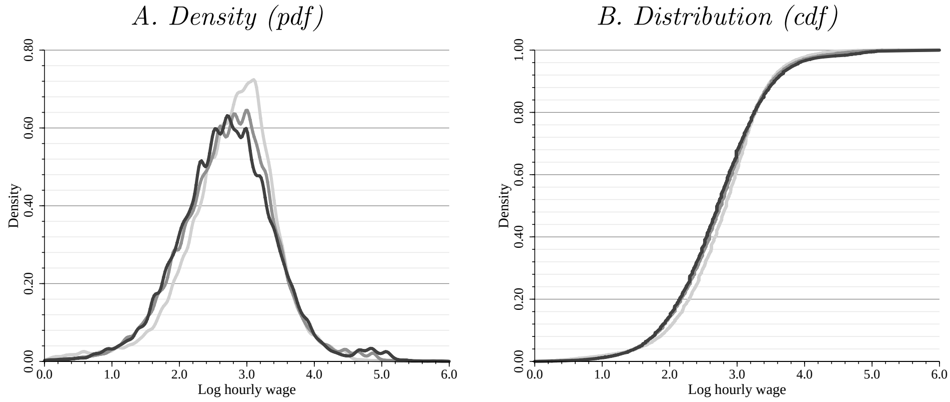
A. Motivation

As in most of the economics literature, so far in the course you have analyzed averages. But sometimes we might be interested in other characteristics of the distribution of our variable of interest, Y_i, beyond the conditional average $\mathbb{E}[Y_i|X_i]$. In particular, in this chapter we are interested in the different *quantiles* of the distribution of Y_i given X_i. As noted below, the τth quantile of the distribution of Y_i is the value q_τ for which a fraction τ of the population has $Y_i \leq q_\tau$. Quantiles fully characterize the distribution of Y_i (and conditional quantiles characterize the distribution of Y_i given X_i), so specifying them is equivalent to specify the cdf of Y_i (or that of Y_i given X_i).

Sometimes, we might not necessarily be interested in the entire distribution, but on a specific quantile. The most popular quantile is the median. We might be interested in the median as a central measure for Y_i as an alternative to the mean. For instance, consider the case in which the data is top-coded. Making inference on the mean is not possible, as it is affected by this *censoring*, but the median might stay unaffected. In that situation, median instead of mean regression might be appealing. The results that are presented in this chapter are general for all quantiles of the distribution, and the median regression can be seen as a special case of this general approach.

An example in which we are interested in non-central characteristics of the distribution of our variable of interest is the study of inequality. For instance, consider the case of wage or income inequality. Figure 1 presents hourly wage distributions (pdfs and cdfs) for U.S. males obtained from Censuses of years 1980, 1990 and 2000. It emerges from the figure that there has been an important increase in inequality during this period. In 1980, there is clearly a larger probability mass in central part of the distribution, whereas in subsequent years part of this mass is distributed over the two tails. Average wages are uninformative of this changing feature of the data (indeed, they have been pretty flat over this period). The

Figure 1. Distribution of U.S. Male Wages (1980-2000)



Note: Light gray: 1980; gray: 1990; dark gray: 2000. Sample restricted to working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. *Data source*: U.S. Census.

tools explained in this chapter allows us to study the determinants, for instance, of changes in wages at different quantiles of the distribution (e.g. why wage at the 20th percentile is 10-20% larger in 1980 than in subsequent years?).[1]

B. *Unconditional Quantiles*

To introduce notation and concepts, it is useful to start with general results for unconditional quantiles. Let $F(Y_i)$ be the cdf of Y_i. The τth quantile of Y_i, $q_\tau(Y_i)$ solves:

$$F(q_\tau(Y_i)) = \tau \qquad \Leftrightarrow \qquad q_\tau(Y_i) = F^{-1}(\tau), \tag{1}$$

or, in words, it is the value of Y_i that leaves a fraction τ of observations below and $1-\tau$ above. Therefore, the set $\{q_\tau(Y_i), \tau \in (0,1)\}$ fully describes the distribution of Y_i. The median, $q_{0.5}(Y_i)$ is the value of Y_i that leaves half of the population above and half below.

The relationship between $q_\tau(Y_i)$ and $F(Y_i)$ described by equation (1) is observed in Figure 1.*ii*. We usually read the plot from the horizontal axis, e.g. in year 2000, $F(2) \approx 0.16$, or, in words, there is a 16% of observations with a log hourly wage below or equal to 2; but $q_\tau(y) = F^{-1}(\tau)$ is equivalent to "reading the plot from the vertical axis", e.g. $q_{0.16} = F^{-1}(0.16) \approx 2.0$, or, in words, the 16th percentile, i.e. the log wage that leaves the 16% of observations below it, is 2.

Table 1 presents estimates for different quantiles of the three wage distributions plotted above. Results from the table are interesting; a sharp increase in inequality

[1] Quantile and percentile are two ways of referring to the same object. For instance, the median is also the 0.5 quantile and the 50th percentile. Other popular terms are quartiles (the 25th, 50th, and 75th percentiles), and deciles (10th, 20th,30th,...).

Table 1—Unconditional Quantiles for Wages (1980-2000)

			Percentile:		
Year	10th	25th	50th	75th	90th
1980	1.96	2.41	2.84	3.18	3.50
1990	1.86	2.30	2.76	3.15	3.51
2000	1.83	2.27	2.70	3.15	3.55

Note: Sample restricted to working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. *Data source*: U.S. Census.

is observed over the three decades (very marked during 1980s). The 10th percentile has decreased by around 13% (around 13 log points), meaning that the individual that, if we order the U.S. population of working men in the year 2000 by their wage, leaves exactly 10% of the population below him earns a wage 13% lower than the guy that leaves 10% of the 1980 U.S. population of working men below him. A similar pattern is observed for the first quartile (the 25th percentile) and for the median. The third quartile (75th percentile) stayed roughly constant, and the top decile (the 90th percentile) increased. Hence, wage inequality increased substantially: individuals at the top of the distribution earn more in 2000 than in 1980, and individuals at the bottom earn less than before.

Sample quantiles There are two ways of computing sample quantiles given a random sample $\{Y_1, ..., Y_N\}$. The first one is to compute the empirical cdf and invert it:

$$\hat{F}_N(r) = \frac{1}{N}\sum_{\text{i=1}}^{\text{N}} \mathbb{1}\{Y_i \leq r\} \qquad \Leftrightarrow \qquad \hat{q}_\tau(Y_i) = \hat{F}_N^{-1}(\tau) \equiv \inf\{r : \hat{F}_N(r) \geq \tau\}. \tag{2}$$

This option is very costly computationally, because it implies ordering all observations and picking the first observation that leaves at least a fraction τ of the sample below it.

The second alternative makes use of the following function, also known as the "check" function, and a nice property of the quantiles.[2] The check function, applied to a given argument u, is:

$$\rho_\tau(u) = \begin{cases} \tau|u| & \text{if } u \geq 0 \\ (1-\tau)|u| & \text{if } u \leq 0 \end{cases}, \tag{3}$$

[2] The nickname of the function comes from its similarity with the check mark ✓.

or, more compactly:

$$\rho_\tau(u) = \tau\, \mathbb{1}\{u \geq 0\}|u| + (1-\tau)\, \mathbb{1}\{u \leq 0\}|u| \equiv \tau u^+ + (1-\tau)u^-. \qquad (4)$$

This function is continuous, but not differentiable at $u = 0$; this later feature prevents the estimators for quantile regression to have a closed form solution as in the standard linear regression model.

A nice property of quantiles that allows us to estimate them in a (computationally) very efficient way is:[3]

$$\hat{q}_\tau(Y_i) = \arg\min_r \sum_{\mathrm{i=1}}^{\mathrm{N}} \rho_\tau(Y_i - r) = \arg\min_r \sum_{Y_i \geq r} \tau|Y_i - r| + \sum_{Y_i \leq r}(1-\tau)|Y_i - r|. \qquad (5)$$

This result is not obvious. To clarify it, consider the first quartile, where $\tau = 0.25$; the first quartile is the minimum of $\sum_{Y_i \geq r} 0.25|Y_i - r| + \sum_{Y_i \leq r} 0.75|Y_i - r|$. Suppose in a sample of 99 observations that the 25th smallest observation (the first quartile), equals 4 and that the 26th observation is equal to 8. If we let r to be 8 instead of 4 (i.e., the 26th observation instead of the 25th), then for the first 25 observations, $Y_i - r$ would be increased by 4 (so the function would increase by $0.75 \times 25 \times 4 = 75$), and for the last 74th, $Y_i - r$ would be decreased by 4 (so the function would decrease by $0.25 \times 74 \times 4 = 74$). Therefore, the 26th observation is a worse candidate than the 25th as a minimizer for the function.

Standard errors Deriving the asymptotic distribution of sample quantiles is out of the scope of this course. The main reason for this is that, given the non-differentiability of the objective function, the asymptotic normality result cannot be established in the standard way. Nonetheless, we can still get an asymptotic normality result using other approaches under suitable conditions, following the more general results for non-smooth GMM estimators. When this is possible, the resulting asymptotic distribution is:

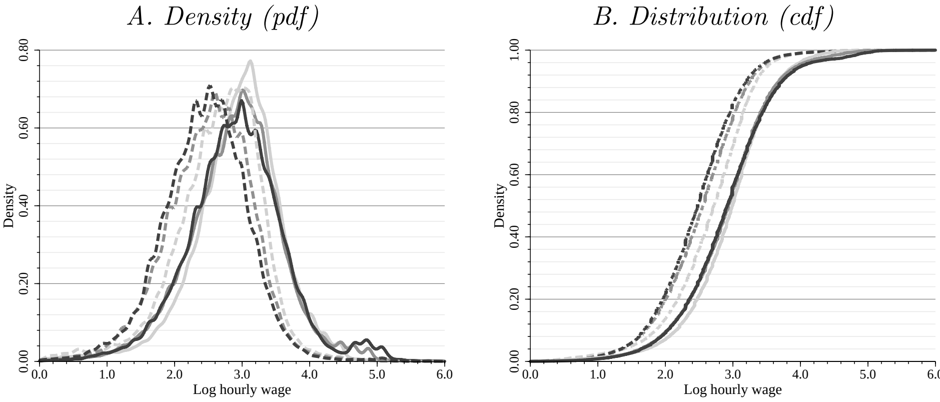
$$\sqrt{N}(\hat{q}_\tau(Y_i) - q_\tau(Y_i)) \underset{d}{\to} \mathcal{N}\left(0, \frac{\tau(1-\tau)}{[f(q_\tau(Y_i))]^2}\right), \qquad (6)$$

where $f(\cdot)$ is the pdf of the distribution $F(\cdot)$.

In practice, standard errors are typically computed using bootstrap (when computationally feasible). However, this asymptotic result is very useful to understand what determines the precision of the estimates. The numerator of the asymptotic variance, $\tau(1-\tau)$, tends to make $\hat{q}_\tau(Y_i)$ more precise in the tails, whereas the density term in the denominator tends to make $\hat{q}_\tau(Y_i)$ less precise in regions of low

[3] This result is the analogue of the population counterpart $q_\tau(Y_i) = \arg\min_r \mathbb{E}[\rho_\tau(Y_i - r)]$.

Figure 2. Distribution of U.S. Male Wages (college vs non-college)



Note: Solid: college; dashed: noncollege. Light gray: 1980; gray: 1990; dark gray: 2000. Sample restricted to working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. Individuals with high school or less are considered as noncollege, whereas individuals with some college or a college degree are considered as college educated. *Data source*: U.S. Census.

density (usually the tails). The latter effect typically dominates so that quantiles closer to the extremes are estimated with less precision.

C. Nonparametric Conditional Quantiles

In this chapter, we are interested in conditional quantiles. The quantile regression model is semiparametric in the sense that it is described by some parameters, but we do not make distributional assumptions. Before we enter into it, we describe a nonparametric example to motivate our interest in conditional quantiles.

Let us revisit our example in Table 1 and Figure 1. Consider the case in which we want to check whether the increasing wage inequality is the result of higher wage earners being increasingly more educated and lower wage earners being less, or, instead, the result holds separately for both college and non-college workers. In the remaining of this section we explore this possibility by comparing empirical distributions for the two groups.

We are, hence, interested in estimating $q_\tau(Y_i|X_i)$, where in this case the vector X_i is indeed a scalar that takes the value of 1 if the individual has college education and 0 otherwise. The nonparametric approach computes sample quantiles for the restricted samples of college and high school workers respectively. Figure 2 plots sample wage distributions for college and non-college educated in 1980, 1990 and 2000. The figure reveals a very different over-time evolution of the wage distribution for the two groups of workers. In particular, the increase in wage inequality has been particularly severe among less educated.

The figure is read in terms of quantiles as follows: college education "increases the 20th percentile" by around 30% (around 30 log points) in 1980, and around 40% (around 40 log points) in 2000, whereas it "increases the 80th percentile" by around 30% in 1980 and around 50% in 2000. In other words, it emerges from the figure that the wage dispersion was similar in 1980 for college and non-college, and it was larger for less educated in 2000.

A relevant point to emphasize about the terminology used in quantile regressions can be seen in the previous paragraph. When we say that "college education increased the 20th percentile by..." we do not mean that the wage of a particular individual if she had college education would be increased by such amount. Instead, the quantile regression coefficients, in the same way that this nonparametric comparison, talk about effects on distributions and not on individuals. Therefore, it means that the particular quantile of the distribution of wages for individuals with college education is larger than the corresponding quantile of the distribution of wages for individuals without college education by such amount.

Although nonparametric comparisons are very appealing when we have a little number of variables to condition upon, each of them with a small amount of points of support, it becomes unfeasible as the number of regressors and/or the number of points of support of each of them grow. This situation, known as the "curse of dimensionality", is common to all nonparametric methods, but it is especially severe in the case of quantiles (e.g. when compared to averages) because we are making inference on the whole distribution instead of on a single moment.

II. Quantile Regression

A. Conditional Quantiles (revisited)

We have seen in the end of the previous section what is the intuition behind conditional quantiles using the nonparametric example. More formally, we can generalize the notation and results shown before for unconditional quantiles to conditional ones.

To define them, we simply have to replace the marginal distribution of Y_i, $F(Y_i)$, for its conditional counterpart $F(Y_i|X_i)$. Given this:

$$q_\tau(Y_i|X_i) = F^{-1}(\tau|X_i). \tag{7}$$

Population conditional quantiles also satisfy:

$$q_\tau(Y_i|X_i) = \arg\min_{q(X_i)} \mathbb{E}[\rho_\tau(Y_i - q(X_i))]. \tag{8}$$

In the nonparametric case, we leave $q_\tau(Y_i|X_i)$ unrestricted, so we have to estimate it for every potential value of X_i. This is what we did before, and it was easy because there was only one regressor which only took two values, but it easily becomes unfeasible as the number of points of support grow.

The quantile regression model, introduced by Koenker and Basset (1978), imposes some parametric structure to $q_\tau(Y_i|X_i))$ in order to allow us to identify it when it is not feasible nonparametrically. In particular, it usually imposes some sort of linear structure in the relationship between Y_i and X_i at the different points of the distribution.

B. The Quantile Regression Model

Location-scale model This is a very simple model that helps us to connect quantile regression methods with classical regression. Consider the following model with conditional heteroskedasticity:

$$Y_i = \mu(X_i;\beta) + \sigma(X_i;\gamma)U_i, \tag{9}$$

where $U_i|X_i \sim G$, independent of X_i. In this model:

$$q_\tau(Y_i|X_i) = \mu(X_i;\beta) + \sigma(X_i;\gamma)G^{-1}(\tau). \tag{10}$$

In this model, all dependence of Y_i on X_i occurs through mean translations —location— (given by $\mu(X_i;\beta)$) and variance re-scaling (given by $\sigma(X_i;\gamma)$):

$$\frac{\partial q_\tau(Y_i|X_i)}{\partial X_i} = \frac{\partial \mu(X_i;\beta)}{\partial X_i} + \frac{\partial \sigma(X_i;\gamma)}{\partial X_i}G^{-1}(\tau). \tag{11}$$

Hence, under homoskedasticity, quantile curves are parallel in this model. Under heteroskedasticity, they are not parallel, but the model is still very restrictive, as the percentage increase in the slope between any pair of quantiles is the same, regardless of the quantiles that are compared.[4]

General quantile regression model A more general quantile regression model is the following:

$$q_\tau(Y_i|X_i) = X_i'\beta_\tau. \tag{12}$$

This model imposes linearity in X_i, but it allows for different effects on different quantiles.

It is important to recall that we have a continuum of quantiles (every individual in the population represents one quantile, and we have infinite individuals in the population). Therefore, in a way, in this model, every individual has a different

[4] More formally, $\partial \ln[q_{\tau_1}(Y_i|X_i) - q_{\tau_2}(Y_i|X_i)]/\partial X_i$ is the same for any pair of quantiles τ_1 and τ_2.

coefficient β_τ. This model, therefore, can be seen as a random coefficients model, where $\beta_\tau = \beta(u)$.

C. Estimation

The estimation of the quantile regression model is analogous to the unconditional estimation of quantiles. In particular, we use the check function, so that:

$$\widehat{\beta}_\tau = \arg\min_b \sum_{\mathrm{i=1}}^{\mathrm{N}} \rho_\tau(Y_i - X_i'b) = \arg\min_b \sum_{Y_i \geq X_i'b} \tau|Y_i - X_i'b| + \sum_{Y_i \leq X_i'b} (1-\tau)|Y_i - X_i'b|. \tag{13}$$

As we saw for the unconditional case, this problem does not have an analytic solution (unlike in OLS), but this minimization problem is (computationally) easy to solve.

A particular case of quantile regression is the median regression. In this case, the check function becomes simply an absolute value, and the estimator is often known as Least Absolute Deviations (LAD) estimator:

$$\widehat{\beta}_{LAD} \equiv \widehat{\beta}_{0.5} = \arg\min_b \sum_{\mathrm{i=1}}^{\mathrm{N}} |Y_i - X_i'b|. \tag{14}$$

This estimator clearly connects with the Least Squares estimator, with the absolute loss replacing the squared one in the minimization problem.

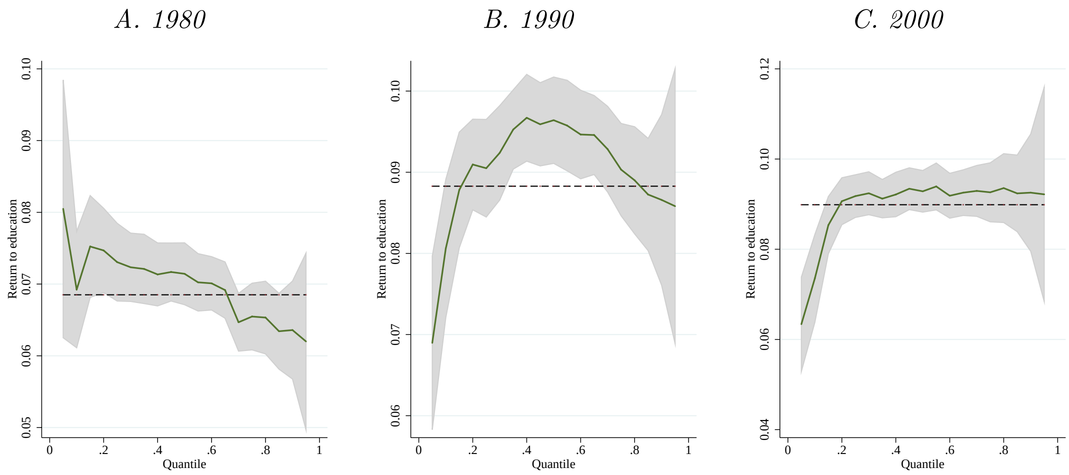
Example Following with our example of wages, consider the following quantile regression model:

$$q_\tau(\ln W_i | E_i, X_i, X_i^2) = \beta_\tau^{(0)} + \beta_\tau^{(E)} E_i + \beta_\tau^{(X)} X_i + \beta_\tau^{(X^2)} X_i^2. \tag{15}$$

Figure 3 plots quantile regression coefficients for education, i.e. $\{\beta_\tau^{(E)} : \tau \in (0,1)\}$. As it emerges from the comparison of the three pictures, returns to education have a different shape for different years, but importantly they have increased all over the distribution except below the bottom quintile (i.e. below the 20th percentile). This increase in the difference in returns to schooling contributed to increase wage inequality.

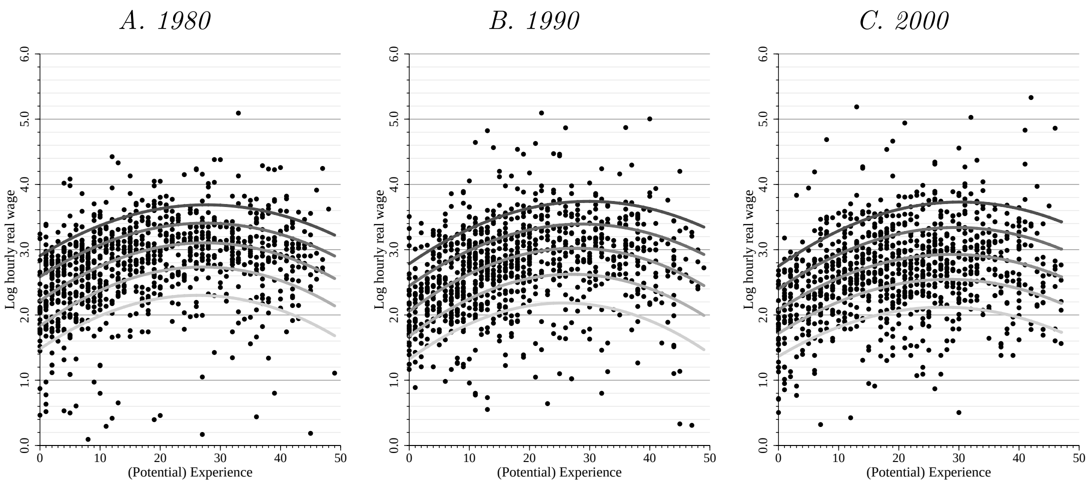
Figure 4 plots the 10th, 25th, 50th, 75th and 90th percentiles of wages conditional on experience. Notice that the quadratic term in experience allows us to have a nonlinear (quadratic in this case) relationship between experience and wages, that seems to be supported by the data at all quantiles. The lines are quite parallel and stable over time, which indicates that experience does not seem to generate much inequality.

Figure 3. Quantile Regression Coefficients (Education)



Note: Random sample of 10,000/year working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. *Data source*: U.S. Census.

Figure 4. Quantiles of Wages Conditional on Experience



Note: Quantiles computed with a random sample of 10,000/year working male aged 16 to 65 who worked at least 20 weeks during the reference year and at least 10 hours per week. Hourly wages are expressed in (log) US$ of year 2000. The scatter plot depicts a random sample of 1,000 observations. *Data source*: U.S. Census.

D. *Quantile Regression with Censoring*

Many times we have *censored* data. In our particular example with wages, it might be that wages are top-coded. If this is the case, mean estimates are consistent anymore, but the median still is (subject to the censoring point being above it). In particular, all quantiles below the censoring point is unaffected by the censoring.

More formally, in the censoring case, if wages are top-coded above a value c, we observe $Y_i^* = \min(y, c)$ instead of y. Then, using an idea by Powell (1986), we can exploit the fact that $q_\tau(Y^*|X_i) = \min(X_i'\beta_\tau, c)$. Hence, we estimate β_τ as:

$$\widehat{\beta}_\tau^c = \arg\min_b \sum_{\mathrm{i=1}}^{\mathrm{N}} \rho_\tau(Y_i - \min(X_i'b, c)). \tag{16}$$

III. Quantile Treatment Effects (QTE)

A. What We Do (and What We Do Not Do)

As in the standard regression case, making causal inference is hard, because regression estimates may be contaminated by omitted variable biases. IV methods for quantile regressions, however, are not so simple, and are still under development these days. When it comes to discrete variables and discrete instruments, however, approaches become simpler. In this context, Abadie, Angrist, and Imbens (2002) introduced the Quantile Treatment Effects (QTE) estimator.

Our conditional quantiles of interest are the following:

$$q_\tau(Y_i|X_i, D_i, D_{1i} > D_{0i}) = \alpha_\tau D_i + X_i'\beta_\tau, \tag{17}$$

where D_i takes the value of one if the individual is treated, and we condition on the fact that the individual is a *complier* ($D_{1i} > D_{0i}$).

In particular, we are interested in α_τ, which is:

$$\alpha_\tau = q_\tau(Y_{1i}|X_i, D_{1i} > D_{0i}) - q_\tau(Y_{0i}|X_i, D_{1i} > D_{0i}). \tag{18}$$

What does this parameter tell us? Consider, as an example, the case in which we want to implement a subsidy for college education (in order to increase college attendance). We want to disentangle what is the effect of increasing college on the distribution of wages. As the decision of attending college is not random, we need an instrument to disentangle what would be the effect of increasing education on wages. In particular, we need an instrument that generates a group of *compliers* that would attend college if it was cheaper but that they do not attend because they find it too expensive. Distance to college might be a good instrument, as the compliers of this group (individuals who do not go if they live far away but go if they live close by) are individuals that would react to the subsidy and change their behavior from not going to going. Hence, our quantile comparison of interest is not between the distribution of wages for individuals who effectively attended college and the one of individuals who did not, but, instead, the distribution of wages of individuals that went to college because it was close to home, but that would have not gone if it had been further away, and that of individuals that did not go because it was far, but would have gone if it had been close. This comparison is what α_τ represents, and what we can identify using this instrument.

However, α_τ does not tell us anything about how the program changed the quantiles of the unconditional distributions of Y_{1i} and Y_{0i}. Also, it does not give us the conditional quantile of the individual treatment effects $q_\tau(Y_{1i} - Y_{0i})$, unlike in the average case: the difference in quantiles is not the quantile of the difference! (although, luckily, economists are typically more interested in the difference in the distributions than in the distribution of the differences)

B. The QTE Estimator

The previous model could be (in theory) estimated consistently in the standard way on the population of compliers. The problem, however, is that we do not observe whether an individual is a complier or not. Alternatively, we can use the Abadie (2003) weighting procedure to compute the appropriate expectation to be minimized.

Let us start with the Abadie (2003) result. If our instrument Z_i satisfies the standard assumptions given X_i, and let $g(Y_i, X_i, D_i)$ be any measurable function of (Y_i, X_i, D_i) with finite expectation, then:

$$\mathbb{E}[g(Y_i, X_i, D_i)|D_{1i} > D_{0i}] = \frac{\mathbb{E}[\kappa_i g(Y_i, X_i, D_i)]}{\mathbb{E}[\kappa_i]}, \tag{19}$$

where:

$$\kappa_i \equiv 1 - \frac{D_i(1 - Z_i)}{1 - \Pr(Z_i = 1|X_i)} - \frac{(1 - D_i)Z_i}{\Pr(Z_i = 1|X_i)}. \tag{20}$$

The main idea is that the operator κ_i "finds compliers". The intuition behind this is that individuals with $D_i(1 - Z_i) = 1$ are *always-takers* as $D_{0i} = 1$ for them; similarly, individuals with $(1 - D_i)Z_i = 1$ are *never-takers*, as $D_{1i} = 0$ for them; hence, if the monotonicity assumption holds, the left-out are the compliers. Indeed:

$$\mathbb{E}[\kappa_i|Y_i, X_i, D_i] = \Pr(D_{1i} > D_{0i}|Y_i, X_i, D_i). \tag{21}$$

Given this result, Abadie, Angrist, and Imbens (2002) developed the QTE estimator as the sample analogue to:

$$(\alpha_\tau, \beta'_\tau) = \arg\min_{(a,b')} \mathbb{E}[\rho_\tau(Y_i - aD_i - X_i'b)|D_{1i} > D_{0i}]$$
$$= \arg\min_{(a,b')} \mathbb{E}[\kappa_i \rho_\tau(Y_i - aD_i - X_i'b)]. \quad (22)$$

Note that the denominator from Abadie's result is irrelevant, as it does not include the parameters of interest.

There are several aspects that worth a mention. First is that κ_i needs to be estimated (and standard errors should take this into account —bootstrapped standard errors, including the estimation of κ_i in the bootstrapping, do). Second, κ_i is negative when $D_i \neq Z_i$ (instead of zero), which makes the regression minimand non-convex. To solve this problem, we can apply the law of iterated expectations, so that we transform the problem into:

$$(\alpha_\tau, \beta'_\tau) = \arg\min_{(a,b')} \mathbb{E}[\mathbb{E}[\kappa_i|Y_i, X_i, D_i]\rho_\tau(Y_i - aD_i - X_i'b)]. \quad (23)$$

This solves the problem as $\mathbb{E}[\kappa_i|Y_i, X_i, D_i] = \Pr(D_{1i} > D_{0i}|Y_i, X_i, D_i)$ is a probability and, hence, is between zero and one. This trick makes indeed the problem very easy to implement in practice. Note that:

$$\mathbb{E}[\kappa_i|Y_i, X_i, D_i] = 1 - \frac{D_i(1 - \mathbb{E}[Z_i|Y_i, X_i, D_i = 1])}{1 - \Pr(Z_i = 1|X_i)} - \frac{(1 - D_i)\,\mathbb{E}[Z_i|Y_i, X_i, D_i = 0]}{\Pr(Z_i = 1|X_i)}. \quad (24)$$

A very simple two-stage method consists of the following two steps:

1) Estimate $\mathbb{E}[Z_i|Y_i, X_i, D_i]$ with a Probit of Z_i on Y_i and X_i separately for $D_i = 0$ and $D_i = 1$ subsamples, and $\Pr(Z_i = 1|X_i)$ with a Probit of Z_i on X_i with the whole sample. Construct $\hat{\mathbb{E}}[\kappa_i|Y_i, X_i, D_i]$ using the fitted values from the previous expressions.[5]

2) Estimate the quantile regression model with the standard procedure (e.g. with `qreg`) using these predicted kappas as weights.

One should then compute the correct standard errors taking into account that the weights are estimated instead of the true weights as we discussed above.

[5] It may happen that, for some observations, the predicted value goes below 0 or above 1; in this case, replace the values below 0 by 0 and the values above 1 by 1.