Conditional vs. Unconditional Quantile Regression Models: A Guide to Practitioners

Javier Alejo[a], Federico Favata[b], Gabriel Montes-Rojas[c,*], Martín Trombetta[d]

[a]IECON - Universidad de la República Uruguay
✉ javier.alejo@ccee.edu.uy

[b]Centro de Investigaciones Macroeconómicas para el Desarrollo - Universidad de Nacional de San Martín
✉ ffavata@unsam.edu.ar

[c]Instituto Interdisciplinario de Economía Política, Universidad de Buenos Aires,
CONICET and Universidad de San Andrés
✉ gabriel.montes@fce.uba.ar   *Corresponding author

[d]CONICET and Universidad Nacional de General Sarmiento
✉ mtrombet@ungs.edu.ar

Abstract

This paper analyzes two econometric tools that are used to evaluate distributional effects, conditional quantile regression (CQR) and unconditional quantile regression (UQR). Our main objective is to shed light on the similarities and differences between these methodologies. An interesting theoretical derivation to connect CQR and UQR is that, for the effect of a continuous covariate, the UQR is a weighted average of the CQR. This imposes clear bounds on the values that UQR coefficients can take and provides a way to detect misspecification. The key here is a match between CQR whose predicted values are the closest to the unconditional quantile. For a binary covariate, however, this relationship is not valid. We illustrate these models using age returns and gender gap in Argentina for 2019 and 2020.

1.   Introduction

In many research areas, it is important to assess the distributional effects of policy variables and key covariates. From a policy maker perspective, an intervention that helps to raise the lower tail of an income or wage distribution is often more appreciated than an intervention that shifts the upper tail of the distribution, even if the average effect is the same. Also an intervention that generates inequality in the distribution of income may be judged inferior to another one that produces less dispersion, even when the average effect may favor the former. This paper analyzes two econometric tools that are used to evaluate these effects, conditional quantile regression (CQR) and unconditional quantile regression (UQR). Our main objective is to shed light on the interpretation of these two methods from the perspective of an applied practitioner, and to focus on the similarities and differences between these two methodologies.

As a starting point, let's specifically define the nomenclature used in the literature (see Fortin et al., 2011). On the one hand, the Conditional Quantile Partial Effect (CQPE) refers to the effect of a covariate (ceteris paribus) on a conditional quantile of an outcome variable, while the same concept on the unconditional distribution is the Unconditional Quantile Partial Effect (UQPE). Conditional here means that we are controlling for the effect of other covariates. On the other hand, CQR and UQR refer to two regression methodologies to estimate CQPE and UQPE, respectively. Sometimes the acronym of the method is informally interchanged with the parameter of interest and therefore can be somewhat confusing if read lightly.

Quantile regression (QR) (see Koenker and Hallock, 2001; Koenker, 2005, for a comprehensive analysis of QR) is a useful way to represent heterogeneity using a set of parameters to characterize the entire conditional distribution of an outcome variable given a list of observable covariates. Let us consider a Mincer equation as a canonical example, i.e. wages as a function of observable workers' characteristics (see Arias et al., 2001, for an illustrative application of QR to this case). The conditional quantiles refer to the salary ranking generated by unobserved characteristics of the individuals (usually ability, motivation, etc.), controlling for the effect that comes from the observables or rather conditional on those observables (experience, educational level, gender, employment sector, etc.). The $\tau \in (0,1)$ index refers to a particular position or rank in the conditional distribution of unobservables, such that τ proportion is below and $1 - \tau$ is above, i.e. for high ability (τ_H) vs. low ability (τ_L). In a simple analysis of the gender gap on wages, QR provides a tool to study heterogeneity in the gender differences across wages in the conditional distribution. Thus the effect at the τth-quantile refers to the wage difference between female and male workers that share the same level of observables covariates and that are located in the same position of the conditional ranking. In another example, say the effect of age on wages, the CQR coefficients estimate the impact of changing from a value x to $x + t$ (where t is small) for the subpopulation that is located in the τ part of the conditional distribution and that also share the same values of other covariates (CQPE). Again we may have a different effect depending on the ranking of unobservables. Note that in this context, x is a particular value of the covariate of interest.

The combination of QR with simulation exercises is usually implemented to evaluate distri-

butional effects. While it is feasible to calculate the unconditional distribution of wages using QR (see Autor et al., 2005; Machado and Mata, 1995; Melly, 2005), this task is not obvious, at least compared to OLS for the conditional mean. Since the law of iterated expectations does not hold in the case of quantiles, the conditional QR analysis cannot be used directly to analyze unconditional quantiles (see the discussion in Fortin et al., 2011). The unconditional quantiles refer to the location in the distribution of the entire wage distribution, making the emphasis in high or low wage values, not necessarily for individuals with the same value in the covariates. It could be the case that those in the η quantile location of the unconditional distribution have certain endowments that makes the comparison with the CQR unclear. In the gender gap example, most high wage earners (let's say η_H) are probably mostly males. Moreover, there are female workers with a high conditional τ_H (talented or motivated workers) but most of them located in the low quantile region (say η_L).

Firpo et al. (2009) propose an implementation of a known statistical tool, the influence function, to evaluate the impact of changes in the covariates on a statistic that depends on the unconditional distribution. Those authors defined that method recentered influence function (RIF) regression where a regression of the RIF on covariates allow to study the marginal effect of a "change" in those covariates. The potential simplicity and flexibility the methodology offers for the analysis of any distributional statistics also motivated subsequent research to expand the use of RIFs in the framework of regression analysis. After its introduction, UQR, i.e. RIF applied to studying the effect of covariates on the unconditional quantiles, became a popular method for analyzing and identifying the distributional effects on outcomes in terms of changes in observed characteristics in areas such as labor economics, income and inequality, health economics, and public policy.

This paper is motivated by the necessity of providing an interpretation of CQPE and UQPE, and also for better understanding the connection between both concepts. We show that the interpretation is as follows. First, for a continuous covariate, the UQPE is determined by a location shift in the entire distribution of a covariate, where the shift corresponds to changing from a random variable X to another $X + t$, where t is small. In the age covariate example, this means evaluating an intervention where the entire labor force is older by one year. Depending on the joint distribution of wages and age, this movement may have an impact that is different at η_L than at η_H. Suppose that a relative higher proportion of conditional high ability (indexed by τ_H) are located in the high η_H part of the unconditional distribution, and that age has a CQR higher impact on high ability than in low ability. Then the UQPE of age on wages should have a larger effect on η_H vis-à-vis η_L. Second, for binary covariates, say D, while the CQPE was clearly understood using regression type tools, for UQPE is not. The key is that the RIF exercise proposes to move the probability of $D = 1$ instead of a particular value D. Then, for the gender gap example, the UQPE should be interpreted as marginally shifting the propensity score of female vs. male, i.e. the proportion of each type. That is, uniformly changing the gender proportion in the sample.

An interesting theoretical derivation to connect CQPE and UQPE is that, for the effect of a continuous covariate, the UQPE is a weighted average of the CQPE. This imposes clear bounds

on the values that UQR coefficients can take and provides a clear connection between CQR and UQR. The key here is a match between CQR whose predicted values are the closest to the unconditional quantile. In fact, the connection of both concepts for the continuous case is useful as a guide to evaluate the correct specification of the regression model. Unfortunately, a similar relationship for a binary covariate does not exist in the literature. We derive a new theoretical result that allows a comparison of the two effects. The analytical results presented here may serve as a guide for understanding the reason of the observed CQR and UQR effects.

This paper is organized as follows. Section 2 discusses CQR models. Section 3 presents UQR. Then Section 4 presents a comparison of this two methods. Section 5 applies this estimators to the study of age and gender effects in wages for Argentina using 2019 and 2020 household surveys. Section 6 concludes. The Appendix A contains mathematical proofs and Appendix B a brief introduction to empirical implementation in STATA and R.

2.   Conditional Quantile Regression

Let Y be the response variable and $\boldsymbol{X}$ be a $p \times 1$ dimensional vector of covariates. The mean and quantile linear regression models are two well known models to estimate the effect of certain covariates on a response variable.

Mean regression (MR) considers the effect of $\boldsymbol{X}$ on Y through the conditional mean model

$$E(Y|\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{\beta}_M, \tag{1}$$

where in this model $\boldsymbol{\beta}_M$ is $p \times 1$ dimensional vector of coefficients.

In quantile regression (QR) the conditional quantiles of Y are of interest through the models

$$Q_Y(\tau|\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{\beta}(\tau) \ \text{ for } \ \tau \in (0,1). \tag{2}$$

In this simple linear model, the parameters $\boldsymbol{\beta}(\tau)$ measure the conditional quantile partial effect (CQPE) of increasing X, that is:

$$CQPE(\tau) = \frac{\partial Q_Y(\tau \mid \boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{\beta}(\tau). \tag{3}$$

Therefore, $\boldsymbol{\beta}(\tau)$ measures the effect of marginally changing X within the group of individuals characterized by the same conditional level of response, indexed by τ. Note that equation (2) implies that the right-hand side is monotone increasing in τ. In theory, the monotonicity requirement should be satisfied for all realizations of X or for some specified subspace of interest, i.e. $\boldsymbol{X} \in \mathcal{X}$ (this is discussed in Koenker, 2005, p. 59). In practice, however, the monotonicity may not be satisfied for some values of $\boldsymbol{X}$, a problem known as the quantile crossing problem: the conditional quantile curves $\boldsymbol{x} \longmapsto Q_Y(\tau|\boldsymbol{x})$ may cross for different values of τ (He, 1997).

Chernozhukov et al. (2009, 2010) studies this monotonicity requirement and proposes a rearrangement procedure of the estimated quantile curves. He (1997) proposed to impose a location-scale regression model, which naturally satisfies monotonicity. Furthermore, Koenker and Xiao (2006) shows that integrating out the quantiles over $(0,1)$ should provide the mean effects, i.e. $\int_{(0,1)} \boldsymbol{\beta}(\tau)d\tau = \boldsymbol{\beta}_M$.

The results in Koenker and Bassett (1982) establish that, under regularity conditions, the estimated conditional quantile function is a strongly consistent estimator of the population quantile function. Thus the process $\{Y, \boldsymbol{X}\}$ can be partially recovered from the marginal distributions, that is, the conditional distribution $Y|\boldsymbol{X}$ can be described by its conditional quantiles based on $\tau \in (0,1)$. The QR analysis constructs a model $Y^* = Y(\boldsymbol{X}, \tau)$ in which Y^*, depends on *endowments* $\boldsymbol{X}$ and its location in the conditional distribution given by τ. The linear QR model determines that the coefficients $\boldsymbol{\beta}(\tau)$ are the *pricings* of those endowments in the market (this is further developed in the next section). Here variation in $\boldsymbol{\beta}(.)$ completely absorbs the role of the unobservables in regression models. This method has been applied for the analysis of inequality by Autor et al. (2005), Machado and Mata (1995), Montes-Rojas et al. (2017), and others.

For a dummy variable (say D), the CQPE should not be obtained by partial derivatives but instead

$$CQPE_D(\tau) = Q_Y(\tau \mid D = 1, \boldsymbol{X}) - Q_Y(\tau \mid D = 0, \boldsymbol{X}). \tag{4}$$

This is the difference in the τ quantiles for the population group with $D = 1$ vis-à-vis those with $D = 0$, in this case controlling for other covariates $\boldsymbol{X}$. This type of estimator is the same used in the *quantile treatment effects* literature (Chernozhukov and Hansen, 2005, 2006; Firpo, 2007), where we compare treated and non-treated were each group is calculated at a given ranking. If the CQR models are linear, then $CQPE_D(\tau) = \beta_D(\tau)$, i.e. the QR coefficient of the dummy variable.

3. Unconditional Quantile Regression

Consider now the case of studying the unconditional quantiles, defined as

$$Q_Y(\eta) \quad \text{for} \quad \eta \in (0,1). \tag{5}$$

We are interested in a multivariate case where we consider the joint distribution of $(Y, \boldsymbol{X})$. In particular, to find a framework to study changes in the distribution of $\boldsymbol{X}$ on the quantiles of Y. This is first a matter of definition, i.e., what is a change in $\boldsymbol{X}$.

A well know result is that quantiles are non-linear operators. As such, while the law of expected iterations can be applied to the conditional mean, this is not valid in general for conditional and unconditional quantiles,

$$Q_Y(\eta) \neq E\left[Q_Y(\eta|\boldsymbol{X})\right] = E[\boldsymbol{X}]'\boldsymbol{\beta}(\eta) \quad \text{for} \quad \eta \in (0,1). \tag{6}$$

Therefore, the CQPE cannot be used directly to study effects in $Q_Y(\eta)$.

Our interest lies in defining the unconditional quantile partial effect (UQPE), but this requires to review influence function theory (see Huber and Ronchetti, 2009 for an introduction).

Consider functionals $v(F)$ defined on the distribution function F_Y of a real valued random variable Y. Typical functionals that have been extensively used in income distribution analysis are inequality, poverty or polarization measures, which are an essential tool to study distributional impacts. Of particular relevance is the use of the influence function (IF) of v, that

summarizes the marginal impact of a particular observation or group in the value of the functional. Moreover it provides a unified approach for computing variances and covariances for general functionals.

The IF is the directional derivative of $v(F_Y)$ at F_Y and it measures the effect of a small perturbation in F_Y. In other words, let y be an additional data point in a large sample that adds a perturbation to the distribution with probability mass δ_y. H_Y is then $H_Y(y) = 1[Y \geq y]$ and $h_Y(y)$ is a density function with value 0 except at y. Then,

$$IF(y; v; F_Y) = \lim_{t \to 0} \frac{v\left[t\delta_y + (1-t)F_Y\right] - v(F_Y)}{t}. \tag{7}$$

The IF of a distributional statistic (and hence of a social evaluation function) measures the relative effect of a small perturbation in the underlying outcome distribution on the statistic of interest. Within that approach and under the assumption that the distributional change in question is due to policy implementation, the IF of a social evaluation criterion may be viewed as a local measure of the distributional impact of policy. For most cases the IF can be computed analytically for a large family of functionals to study inequality and poverty (such as quantiles, variance, Gini, etc.).

The IF is a key tool to study the distributional impact of changes in covariates. The re-centered influence function (RIF) is defined as

$$RIF(Y, v, F_Y) = v(F_Y) + IF(Y, v, F_Y), \tag{8}$$

which is motivated by the fact that $E_Y[IF(Y, v, F_Y)] = 0$, and then

$$v(F_Y) = E_Y[RIF(Y, v, F_Y)]. \tag{9}$$

For the particular case of the η-quantile, $v(F_Y) = Q_Y(\eta) = F_Y^{-1}(\eta)$, we have (see Essama-Nssah and Lambert, 2015 for a derivation)

$$RIF(y; Q_Y(\eta), F_Y) = Q_Y(\eta) + \frac{\eta - 1[y \leq Q_Y(\eta)]}{f_Y(Q_Y(\eta))}. \tag{10}$$

Note that property of the RIF in eq. (9) is very interesting because it allows to compute any indicator $v(F_Y)$ as an expectation and therefore enables us to use the Law of Iterated Expectations

$$v(F_Y) = E\left\{E\left[RIF(Y, v, F_Y) \mid \boldsymbol{X}\right]\right\}. \tag{11}$$

The conditional expression $E\left[RIF(Y, v, F_Y) \mid \boldsymbol{X} = \boldsymbol{x}\right]$ is an unknown function $m(\boldsymbol{x})$, then $v(F_Y) = E[m(\boldsymbol{X})]$.

Now suppose an small location-shift of the distribution of one covariate X, that is $X + t$ where t is small. For simplicity, consider a single covariate, but the analysis should be interpreted as a change in one covariate keeping the others unaltered. This shift affects the entire unconditional distribution of Y, moving it towards a new distribution G_Y. Note here the mechanism: the distribution of X is shifted, which has an impact on the joint distribution of (Y, X), for which we want the effect on the distribution of Y. Compute the marginal change from $v(F_Y)$ to $v(G_Y)$:

$$\lim_{t \to 0} \frac{v(G_Y) - v(F_Y)}{t} = \lim_{t \to 0} \frac{E[m(X+t)] - E[m(X)]}{t} = E\left[\lim_{t \to 0} \frac{m(X+t) - m(X)}{t}\right],$$

where the last equality uses the continuity of the limit.

If $m(x)$ is differentiable at any point x, then

$$\lim_{t \to 0} \frac{v(G_Y) - v(F_Y)}{t} = E\left[\frac{\partial m(x)}{\partial x}\right] = E\left[\frac{\partial E\left[RIF(Y, v, F_Y) \mid X = x\right]}{\partial x}\right].$$

This is the main equation in Firpo et al. (2009), which is an application of the IF theory and the Law of Iterated Expectations.

Consider now the example of a dummy variable D. In this case, the UQPE and CQPE cannot be compared using the formulas above because we need to consider discrete change. To simplify notation we only consider a model $m(D, X)$ for the conditional expectation of the RIF, where D is a binary covariate and X is a continuous covariate. Then

$$v(F_Y) = E[m(D, X)] = E[m(1, X)P_F(D = 1, X) + m(0, X)P_F(D = 0, X)]. \tag{12}$$

Then consider the following shift:

$$P_G(D = 1, x) = P_F(D = 1, x) + t \tag{13}$$

and

$$v(G_Y) = E[m(D, X)] = E[m(1, X)P_G(D = 1, X) + m(0, X)P_G(D = 0, X)]. \tag{14}$$

Note that when t is small the change in eq. (13) is close to $P_G(D = 1) = P_F(D = 1) + t$.

Compute the difference and use some linear properties of the expectation

$$\lim_{t \to 0} \frac{v(G_Y) - v(F_Y)}{t}$$

$$= \lim_{t \to 0} \frac{E\{m(1, X)[P_G(D = 1, X) - P_H(D = 1, X)] + m(0, X)[P_G(D = 0, X) - P_H(D = 0, X)]\}}{t}$$

Note that $P_G(D = 1, X) - P_H(D = 1, X) = t$ and $P_G(D = 0, X) - P_H(D = 0, X) = -t$, then

$$\lim_{t \to 0} \frac{v(G_Y) - v(F_Y)}{t} = \lim_{t \to 0} \frac{E\{m(1, X)t + m(0, X)(-t)\}}{t}.$$

Finally,

$$\lim_{t \to 0} \frac{v(G_Y) - v(F_Y)}{t} = E\{m(1, X) - m(0, X)\} = E\{\Delta_D E[RIF(Y, v, F_Y)|X]\},$$

where

$$\Delta_D E[RIF(Y, v, F_Y)|X] \equiv E[RIF(Y, v, F_Y)|D = 1, X] - E[RIF(Y, v, F_Y)|D = 0, X], \tag{15}$$

is the partial difference taken on the conditional expectation of the RIF.

When $v(F_Y) = Q_Y(\eta)$, the above procedure is defined as Unconditional Quantile Regression (UQR) and thus,

$$UQPE_X(\eta) = E\left\{\frac{\partial E[RIF(Y, Q_Y(\eta), F_Y) \mid X = x]}{\partial x}\right\} \tag{16}$$

for the continuous case and

$$UQPE_D(\eta) = E\left\{E\left\{\Delta_D E\left[RIF(Y, v, F_Y)|X\right]\right\}\right\}, \tag{17}$$

for the dummy variable case.

Firpo et al. (2009) develop a simple regression framework that is similar to a standard regression except that the dependent variable, Y, is replaced by the re-centered influence function (RIF) of the statistic of interest. They propose $m(\boldsymbol{X}) = \boldsymbol{X'b}$ as an approximation to the functional form of the RIF conditional expectation. The parameters $\boldsymbol{b}$ correspond to a change in the covariates on v that can be interpreted as above. In particular, as explained in Fortin et al. (2011), the *RIF-regression* allows for using regression type tools, such as Oaxaca-Blinder type decompositions. Note however, that this is only an approximation to a potentially nonlinear function, $m(\boldsymbol{X})$, and as such it may fail to appropriately describe the marginal effects.

4. Comparison of UQR and CQR

We have established that QR measures the partial effects of moving $\boldsymbol{X}$ on the conditional distribution (CQPE), while RIF regressions do the same with the unconditional distribution (UQPE). Proposition 1 in Firpo et al. (2009) is very useful to understand the relationship between both statistics. Define the following matching function:

$$\xi_\eta(\boldsymbol{x}) \equiv \left\{\tau : Q_Y[\tau|\boldsymbol{X} = \boldsymbol{x}] = Q_Y(\eta)\right\}.$$

This function finds the τ quantile in the CQPE model that is the "closest" to the η quantile in the UQPE model. It plays a central role in the following analysis.

4.1 Continuous Covariate Case

Consider first the case of a continuous covariate. Figure 1 shows a very simple example of how $\xi_\eta(\boldsymbol{x})$ works for the case of a one dimensional covariate. Suppose we are positioned at the median
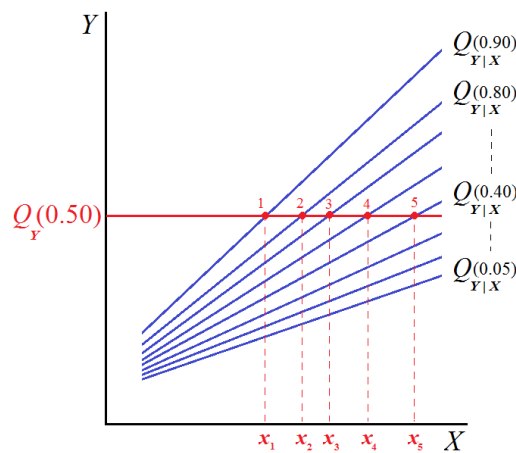

Figure 1. Matching function.

Table 1

Matching function.

Intersection	η	X	$\xi_\eta(X)$
1	0.50	x_1	0.90
2	0.50	x_2	0.80
3	0.50	x_3	.
4	0.50	x_4	.
5	0.50	x_5	0.40

of the unconditional distribution, that is, at $\eta = 0.50$. This value matches with the quantile 0.90 conditional on $X = x_1$, but also with the quantile 0.80 conditional on $X = x_2$, and so on. Then, the function $\xi_\eta(x)$ indicates intersections such as points 1, 2, 3, ... generating a correspondence as shown in Table 1.

A relationship between UQPE and CQPE can be established using the following equation (Proposition 1 (ii) in Firpo et al., 2009):

$$UQPE(\eta) = E\left[\omega_\eta(\boldsymbol{X}) \cdot CQPE\left(\xi_\eta(\boldsymbol{X}), \boldsymbol{X}\right)\right], \tag{18}$$

where the weight has the following form $\omega_\eta(\boldsymbol{X}) \equiv \frac{f_{Y|\boldsymbol{X}}(Q_Y(\eta)|\boldsymbol{X})}{f_Y(Q_Y(\eta))}$ and we allow the CQPE to depend on the value of the covariates. That is, the $UQPE(\eta)$ is a weighted average of all the $CQPE(\tau, \boldsymbol{x})$ that are intersected by $Q_Y(\eta)$. Following the example of Figure 1, the UQPE is the average of the slopes of points 1, 2, 3,... weighted by $\omega_\eta(x)$. If we make this calculation more explicit we can see all the factors that link UQPE with CQPE:

$$UQPE(\eta) = \int \omega_\eta(\boldsymbol{x}) \cdot \text{CQPE}\left(\xi_\eta(\boldsymbol{x}), \boldsymbol{x}\right) \cdot f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}.$$

Note that in fact, the comparison of the two models is similar to the one used in binary regression models (i.e. probit or logit) to get the average marginal effects on the probability. This is because we need to calculate the conditional expectation of $E[1[y \leq Q_Y(\eta)]|\boldsymbol{X}] = P[y \leq Q_Y(\eta)|\boldsymbol{X}]$. When we compute the marginal effect, the derivative of the cumulated probability becomes the density function (i.e. $f_{Y|\boldsymbol{X}}(Q_Y(\eta)|\boldsymbol{X})$) multiplied by the corresponding beta coefficient (i.e. CQPE).

Although both the marginal and joint distributions of $(Y, \boldsymbol{X})$ intervene in the weighting factor, the heterogeneity of the weights is affected by $f_{Y|\boldsymbol{X}}(.)$ and $f_{\boldsymbol{X}}(.)$. Let's consider two simple examples with a location-shift model where:

i) $Q_Y[\tau|\boldsymbol{X} = \boldsymbol{x}]$ is linear and the QR coefficients are constant across quantiles as in Figure 2, then trivially $UQPE(\eta) = CQPE(\eta)$ for any weighting structure;

ii) $Q_Y[s|\boldsymbol{X} = \boldsymbol{X}]$ is non-linear and quadratic as in Figure 3 with $X \sim Uniform$ and $Y|X \sim Uniform$ (such that weights are constant), then it is easy to see that $CQPE(\eta) \neq UQPE(\eta) = 0$ for all η since the partial effects of the conditional distribution are offset in the tails of the distribution of $\boldsymbol{X}$, dissappearing the effects on the unconditional distribution.
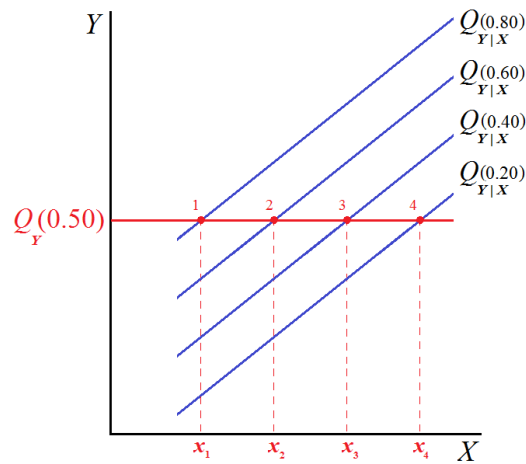
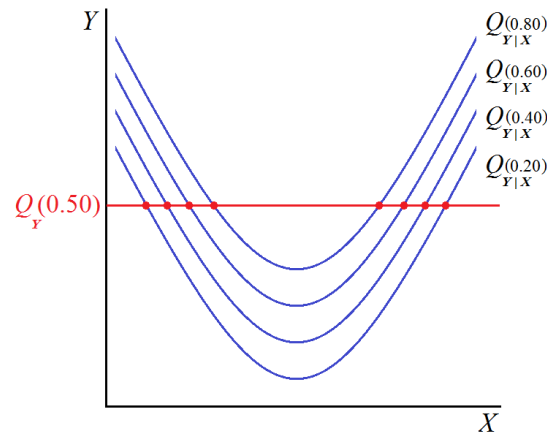**Figure 2.** Linear location-shift model.

Figure 3. Nonlinear quadratic location-shift model.

In practice, all these calculations are not really necessary since we can directly estimate $CQPE(\eta)$ and $UQPE(\eta)$ separately using CQR and UQR-RIF-regression, respectively. However, the relationship discussed in this section is useful to understand the origin of the differences between the two estimates.

It also suggests a way to evaluate model misspecification. In particular, if the UQPE heterogeneity exceeds that of CQPE, that is, if the effects from UQPE exceeds (in absolute values) that of CQPE, then the models are misspecified. This is a relevant issue as quantile and RIF regression are usually implemented as linear models, which could be a poor approximation to the true data generating process.

4.2 Binary Discrete Covariate Case

For the binary discrete covariate case, the function above linking UQR and CQR does not apply. However, a comparison of the eqs. (4) and (17) can be studied in the following proposition.

Proposition 1 *Consider a model that only has one dummy covariate D. Then, we have the following representation*

$$UQPE(\eta) = E\left\{\frac{f_{Y|D}[Q_Y(\eta)]}{f_Y(Q_Y(\eta))}CQPE[\xi_\eta(1-D)]\right\} + R(\tilde{\xi}_\eta),$$

with $R(\tilde{\xi}_\eta)$ defined in the Appendix.

Proof: In the Appendix A.

This result determines that it is not possible to bound UQR effects within those of CQR as in the continuous case. That is, if we frame the case of a dummy covariate for UQR, the UQPE formula for the continuous case is incorrect. In turn, the remainder term $R(\tilde{\xi}_\eta)$ depends on the comparison of the corresponding τ conditional quantiles coming from the matching function. That is, $R(\tilde{\xi}_\eta)$ depends on the difference $\xi_\eta(1) - \xi_\eta(0)$, and since $\xi_\eta(1)$ and $\xi_\eta(0)$ are discrete values, nothing guarantees that it is close to zero.

5.   Empirical Application

We use microdata from the Argentine household survey, *Encuesta Permanente de Hogares* (EPH). This survey is carried out by the National Statistics Institute (INDEC) and collects data from 31 major urban areas, with a statistical coverage rate of approximately 62% of total population. The sampling process involves two stages based on geographical stratification—each one of the 31 cities is first divided into census radii, from which households are randomly selected. The survey's rotating scheme allows for the construction of short-term panels but we have not explored that feature in this application.

Databases are released quarterly and we will use 2019q4 and 2020q4. Note that the surveys correspond to one pre- and another post-COVID-19 pandemic. This choice stems from the fact that social confinement measures were particularly strict in most Argentine cities and we intend to check whether this may have influenced our results. Sample size is 19,662 for 2019 and 12,499 for 2020. We include all salaried or self-employed individuals aged 18 to 65.

We estimate a Mincer regression in order to illustrate the differences among conditional quantile regression (CQR), unconditional quantile regression (UQR) and ordinary least squares (OLS). In particular, we compute (i) the marginal effect of one year of age (continuous variable) on log hourly wages, and (ii) the gender log hourly wage gap between men and women (dummy variable). For the former, we calculate the marginal effect for an individual with average sample age using the delta method to get the standard errors. Formally, the estimated equation is

$$\ln wage_i = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 female_i + x_i'\gamma + u_i,$$

where x is a vector that includes other covariates—education, region, citizenship, economic sector, seniority, firm size, hours and position.

Figure 4 shows the empirical results for the effect of age. OLS estimates are of course independent of the quantile considered, so that one additional year of age always raises the mean
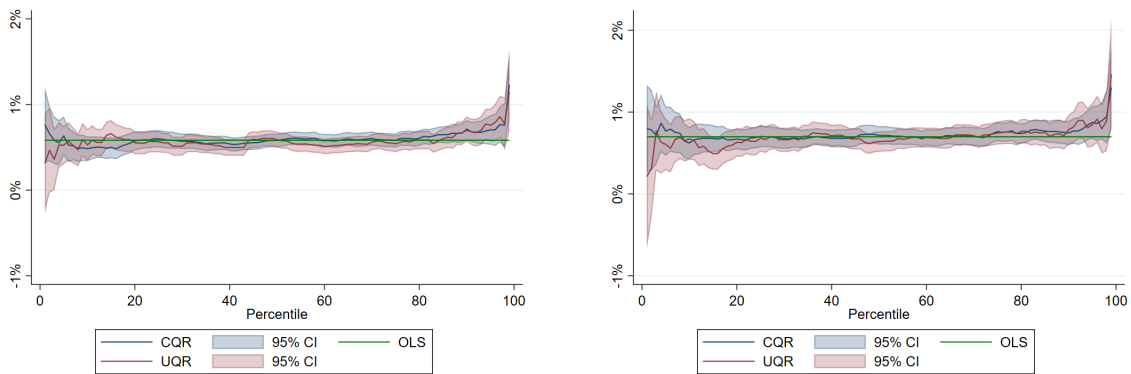
(a) Age - 2019q4.

(b) Age - 2020q4.

Figure 4. Marginal effects: OLS, CQR and UQR, Age.

individual's wage in about 0.6%, no matter whether the individual has low or high (conditional or unconditional) wage. The CQR and UQR are somewhat different until percentile 30, where the distance between them becomes smaller. Both quantile effects are close to 1% for the highest quantiles. However, OLS seems a reasonable approximation for the effect in essentially any point of the distribution, whether conditional or unconditional.

Figure 5 shows that the gap between wages of females and males is about -14%. Here the 2019 curve shows an inverted U shape pattern that is not as clear in the 2020 curve. The gender gap is particularly large in the left tail, where the CQR effect is below -20% and the UQR can reach -50%. Both CQR and UQR are located somewhat above the OLS level (around -11%) between percentiles 20 and 80, but behavior at the right tail differs between the yearly samples and therefore is more difficult to interpret.

Let us illustrate the important differences between CQR and UQR by elaborating on the latter finding. A linear model of the wage equation indicates that the gender wage gap is about 14%, but CQR shows that it can soar to 20% for the lowest conditional wages. This means that
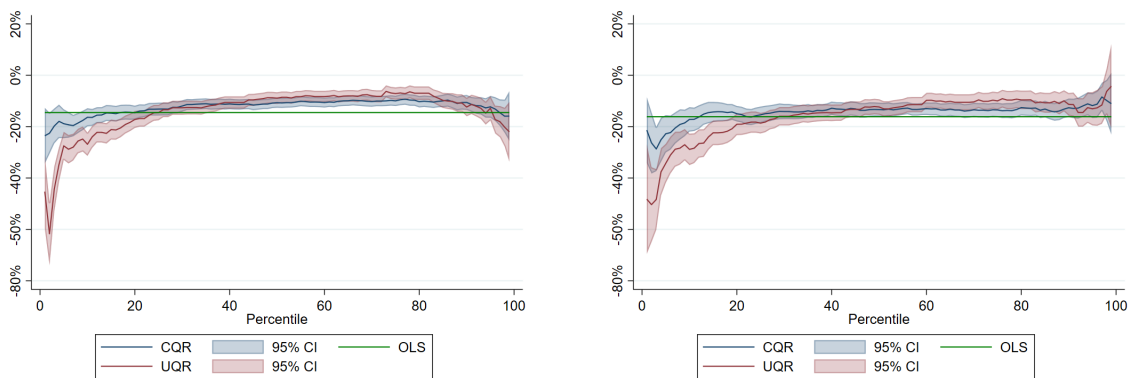
(a) Gender - 2019q4.

(b) Gender - 2020q4.

Figure 5. Marginal effects: OLS, CQR and UQR, Gender Gap.

individuals who have unusually low wages given their observables show a larger gender gap than those whose wages are closer to the mean. But it does not say anything about whether this is true for individuals whose wage is low independently of their observables. UQR shows that, not only this is also the case, but the gender gap for those individuals can be up to 50%. This determines that gender differences are not only observed after controlling for human capital variables, but the latter are also associated with gender differences in such a way that the unconditional effect is much larger.

The fact that confidence intervals overlap sometimes implies that some of these differences might not be statistically significant. This is partly a consequence of limited sample sizes. This explains why confidence intervals are narrower in the 2019 sample (that contains almost 50% more observations), thus making some differences more salient—notably, the left tail of the distribution of the gender gap, where the intervals barely overlap.

We can observe how the results of the empirical application are in agreement with the theoretical explanations. Thus, the marginal effects of the age (continuous variable) of the unconditional regressions (UQPE) are contained within the marginal effect of the conditional regressions (CQPE). However, it can be observed that for gender (discrete variable), particularly in the lower quantiles of the distribution, the marginal effects of the UQR is not contained within the conditional effect. This is because it is not possible to obtain a clear theoretical relationship between UQPE and CQPE for discrete variables.

To gain some intuition on the previous empirical results, we introduce the following CQ-UQ curve. For each value of individual i we calculate $\tau_i^* = \min_\tau \{Y_i \le Q_Y(\tau|\boldsymbol{x}_i)\}$, that is, the corresponding conditional quantile τ that matches closely the individual's wage. Then we estimate for all η, $\tau(\eta) = \Pr(Y \le Q_Y(\tau^*|\boldsymbol{x})|Y \le Q_Y(\eta))$. Then we consider the graph $(\eta, \tau(\eta))$ that matches the unconditional quantiles η with the estimated conditional distribution in order to study if there is any relationship between the two. Figure 6 shows the results. We find that individuals with low values of τ are largely over-represented among those with low values of η,
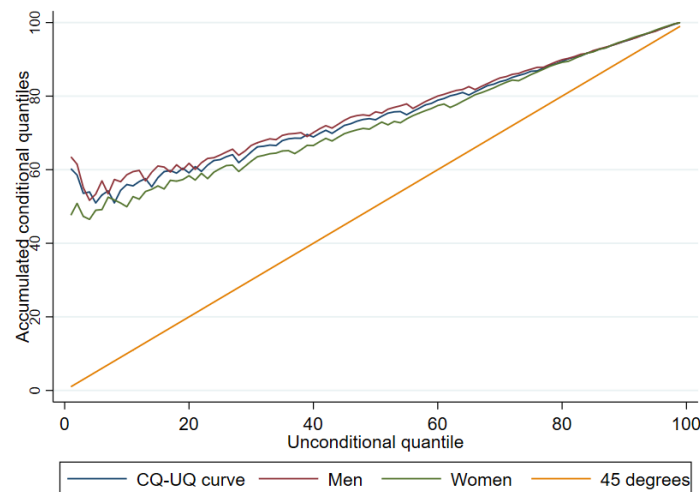

Figure 6. CQ-UQ, Gender Gap.

implying that the left tail of the income distribution consists mostly of individuals whose income level is low conditional on their observable attributes.

Furthermore, the CQPE-UQPE curve for men lies above the one for women, therefore the over-representation of low τ cases in the left tail of the unconditional distribution is higher for men. This could help rationalize the fact that the gender gap is larger in the low unconditional quantiles but not so much so in the low conditional quantiles.

6.   Conclusion

This paper is motivated by the difficulty in understanding the use of two increasingly popular methods to estimate distributional impacts, conditional and unconditional quantile regression. We focus on their interpretation and establish the connection between both models from the point of view of an applied practitioner.

We study the similarities and differences of these two. In particular, we show that for the case of continuous covariates there is a clear link between these two, that is, one is a weighted average of the other. For non-continuous or binary covariates we show that the previous result does not apply, and we establish a new analytical relationship.

The present paper can be extended along several directions. First, the linear approximation of the RIF model can be studied to establish when it works and when it does not. In this case, exploring non-linear models for CQR and UQR may illustrates their potential pitfalls. Second, the analytical connection between the two motivates new estimators where CQR is estimated first, and then UQR is derived. This avoids using the RIF linear approximation. Finally, further empirical statistics based on the comparison between both methods can be informative about population characteristics.

Appendix A - Proof of Proposition 1.

Consider a model that only has one dummy covariate D. Using the results above for this case case the UQPE is:

$$UQPE(\eta) = E\left[RIF(Y; Q_Y(\eta), F_Y)|D = 1\right] - E\left[RIF(Y; Q_Y(\eta), F_Y)|D = 0\right].$$

Namely,

$$UQPE(\eta) = -\frac{Pr[y \leq Q_Y(\eta)|D = 1] - Pr[y \leq Q_Y(\eta)|D = 0]}{f_Y(Q_Y(\eta))}. \tag{19}$$

From the matching function definition it is easy to see that:

$$\xi_\eta(d) = Pr[Y \leq Q_Y(\eta)|D = d] = F_{Y|D=d}[Q_Y(\eta)], \tag{20}$$

such that

$$Q_Y(\eta) = Q_{Y|D=0}[\xi_\eta(0)] = Q_{Y|D=1}[\xi_\eta(1)].$$

Thus,

$$UQPE(\eta) = -\frac{\xi_\eta(1) - \xi_\eta(0)}{f_Y(Q_Y(\eta))}. \tag{21}$$

In addition, also from the definition of matching function:

$$Q_{Y|D=1}[\xi_\eta(1)] - Q_{Y|D=0}[\xi_\eta(0)] = 0$$

$$Q_{Y|D=1}[\xi_\eta(1)] - Q_{Y|D=0}[\xi_\eta(0)] \pm Q_{Y|D=d}[\xi_\eta(1-d)] = 0$$

$$Q_{Y|D=1}[\xi_\eta(1)] - Q_{Y|D=d}[\xi_\eta(1-d)] + Q_{Y|D=d}[\xi_\eta(1-d)] - Q_{Y|D=0}[\xi_\eta(0)] = 0$$

Note that there are two cases: (i) for $d = 0$

$$CQPE[\xi_\eta(1)] + Q_{Y|D=0}[\xi_\eta(1)] - Q_{Y|D=0}[\xi_\eta(0)] = 0, \tag{22}$$

and (ii) for $d = 1$

$$Q_{Y|D=1}[\xi_\eta(1)] - Q_{Y|D=1}[\xi_\eta(0)] + CQPE[\xi_\eta(0)] = 0. \tag{23}$$

Let us then consider the following expansion of $Q_{Y|D=d}[\xi_\eta(1-d)]$ about $\xi_\eta(d)$,

$$Q_{Y|D=d}[\xi_\eta(1-d)] = Q_{Y|D=d}[\xi_\eta(d)] + \frac{\partial Q_{Y|D=d}[\xi_\eta(d)]}{\partial \tau}[\xi_\eta(1-d) - \xi_\eta(d)] + R(\tilde{\xi}_\eta, d),$$

with $R(\tilde{\xi}_\eta, d) = \frac{1}{2}\frac{\partial^2 Q_{Y|D=d}[\tilde{\xi}_\eta]}{\partial \tau^2}[\xi_\eta(1-d) - \xi_\eta(d)]^2$, where $\tilde{\xi}_\eta$ is some value between $\xi_\eta(d)$ and $\xi_\eta(1-d)$.

Also, we know that $\partial Q_{Y|D}(\tau)/\partial \tau = 1/f_{Y|D}[Q_{Y|D}(\tau)]$ and note also that $f_{Y|D}\{Q_{Y|D=d}[\xi_\eta(d)]\} = f_{Y|D=d}[Q_Y(\eta)]$.

Substituting everything in (22) and (23) we can solve for two expressions for $\xi_\eta(1) - \xi_\eta(0)$:

$$\xi_\eta(1) - \xi_\eta(0) = -f_{Y|D=0}[Q_Y(\eta)]\{CQPE[\xi_\eta(1)] - R(\tilde{\xi}_\eta, 0)\}$$

for the case $d = 0$ and

$$\xi_\eta(1) - \xi_\eta(0) = -f_{Y|D=1}[Q_Y(\eta)]\{CQPE[\xi_\eta(0)] + R(\tilde{\xi}_\eta, 1)\}$$

when $d = 1$.

Averaging both cases (weighted by their probability), substituting it in (21) and rearranging terms, we have

$$UQPE(\eta) = \frac{f_{Y|X=0}[Q_Y(\eta)]}{f_Y(Q_Y(\eta))}CQPE[\xi_\eta(1)]Pr(X = 0)$$
$$+ \frac{f_{Y|X=1}[Q_Y(\eta)]}{f_Y(Q_Y(\eta))}CQPE[\xi_\eta(0)]Pr(D = 1) + R(\tilde{\xi}_\eta),$$

where

$$R(\tilde{\xi}_\eta) = \frac{f_{Y|D=0}[Q_Y(\eta)]}{f_Y(Q_Y(\eta))}R(\tilde{\xi}_\eta, 0)Pr(D = 0) + \frac{f_{Y|X=1}[Q_Y(\eta)]}{f_Y(Q_Y(\eta))}R(\tilde{\xi}_\eta, 1)Pr(D = 1).$$

Then,

$$UQPE(\eta) = E\left\{\frac{f_{Y|D}[Q_Y(\eta)]}{f_Y(Q_Y(\eta))}CQPE[\xi_\eta(1 - D)]\right\} + R(\tilde{\xi}_\eta),$$

where $R(\tilde{\xi}_\eta) = E\left\{\frac{f_{Y|D}[Q_Y(\eta)]}{f_Y(Q_Y(\eta))}R(\tilde{\xi}_\eta, D)\right\}$.

Appendix B - Guide for Implementation in STATA and R

To apply conditional quantile regressions in STATA, use the `qreg` command. Standard syntax is `qreg depvar indepvars, q(`τ`)`. See https://www.stata.com/manuals/rqreg.pdf.

Similarly, the `rq` command, from the `R quantreg` library, can be used. Standard syntax for this function is `rq(depvar ~ indepvars, tau = `τ`)`. See https://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf.

To apply unconditional quantile regressions in STATA, you must use the `rifhdreg` command. The syntax is `rifhdreg depvar [indepvars], rif(q(`τ`))`. You will have to install the command through the following syntax: *ssc install rifhdreg* and then obtain the manual by `help rifhdreg` (Rios-Avila, 2020).

Likewise, to apply unconditional quantile regression in R, install the library "uqr". The standard syntax is `urq(depvar ~ indepvars, tau = `τ`)`. UQR code is available at http://cran.nexr.com/web/packages/uqr/index.html.

References

Arias, O., Hallock, K., and Sosa-Escudero, W. (2001). Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data. *Empirical Economics 26*, 7–40.

Autor, D. H., Katz, L. S., and Kearney, M. S. (2005). Rising Wage Inequality: The Role of Composition and Prices. NBER Working Paper No. 11628.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika 96 (3)*, 559–575.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and Probability Curves Without Crossing. *Econometrica 78 (3)*, 1093–1125.

Chernozhukov, V., and Hansen, C. (2005). An IV Model of Quantile Treatment Effects. *Econometrica 73 (1)*, 245–261.

Chernozhukov, V., and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effects models. *Journal of Econometrics 132 (2)*, 491–525.

Essama-Nssah, B., and Lambert, P. J. (2015). Influence Functions for Policy Impact Analysis. In J. A. Bishop and R. Salas (Eds.), *Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber (Research on Economic Inequality, vol. 20)* (pp. 135–159). Bingley: Emerald Group Publishing Limited.

Firpo, S. (2007). Efficient Semiparametric Estimation of Quantile Treatment Effects. *Econometrica 75 (1)*, 259–276.

Firpo, S., Fortin, N., and Lemieux, T. (2009). Unconditional Quantile Regression. *Econometrica 77 (3)*, 953–973.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition Methods in Economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, vol. 4 (pp. 1–12). Amsterdam: Elsevier.

He, X. (1997). Quantile Curves Without Crossing. *The American Statistician 51 (2)*, 186–192.

Huber, P., and Ronchetti, E. (2009). *Robust Statistics*. Cambridge: Wiley.

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

Koenker, R., and Bassett, G. W. (1982). Tests of Linear Hypothesis and l_1 Estimation. *Econometrica 50 (6)*, 1577–1584.

Koenker, R., and Hallock, K. (2001). Quantile Regression. *Journal of Economic Perspecives 15 (4)*, 143–156.

Koenker, R., and Xiao, Z. (2006). Quantile Autoregression. *Journal of the American Statistical Association 101 (475)*, 980–990.

Machado, J. A. F., and Mata, J. (1995). Counterfactual Decomposition of Changes in Wage Distributions Using Quantile. *Journal of Applied Econometrics 20 (4)*, 445–465.

Melly, B. (2005). Decomposition of differences in distribution using quantile regressions. *Labour Economics 12 (4)*, 577–590.

Montes-Rojas, G., Siga, L., and Mainali, R. (2017). Mean and quantile regression Oaxaca-Blinder decompositions with an application to caste discrimination. *Journal of Economic Inequality 15*, 245–255.

Rios-Avila, F. (2020). Recentered influence functions (RIFs) in Stata: RIF regression and RIF decomposition. *The Stata Journal 20(1)*, 51–94.