Unconditional Quantile Regressions *

Sergio Firpo — PUC-Rio and UBC

Nicole Fortin — UBC

Thomas Lemieux — UBC

June 20 2006

Preliminary Paper, Comments Welcome

Abstract

We propose a new regression method for modelling unconditional quantiles of an outcome variable as a function of explanatory variables. The method consists of running a regression of the (recentered) influence function of the unconditional quantile of the dependent variable on the explanatory variables. The influence function is a widely used tool in robust estimation that can easily be computed for each quantile of interest. The estimated regression model can be used to infer the impact of various explanatory variable on a given unconditional quantile, just like the regression coefficients are used in the case of the mean. Our approach can thus be used, for example, to decompose quantiles as a function of the different explanatory variables (as in a standard Oaxaca-Blinder mean decomposition), or to predict the effect of changes in policy or other variables on quantiles.

Keywords: *Influence Functions, Quantile Regressions.*

*We are indebted to Richard Blundell, David Card, Geert Ridder, Jean-Marc Robin, Hal White and seminar participants at the CESG 2005 Meetings for useful comments on this and earlier versions of the manuscript. We thank Kevin Hallock for kindly providing the birthweight data used in the application. Financial support was provided by SSHRC Grants INE #512-2002-1005.

1 Introduction

Quantile regressions are an increasingly popular method in empirical economics. Just like OLS regressions provides a simple way of estimating the effect of explanatory variables on the conditional mean of an outcome variable, quantile regressions provide simple estimates of the effect of the same variables on any conditional quantile of the outcome variable. Quantile regressions enable researchers to "go beyond the conditional mean", which is useful for a number of reasons. In particular, in many cases we not only care about the mean of the outcome variable, but also about other aspects of the distribution. For instance, Koenker and Hallock (2001) apply quantile regressions to the case of infant birthweight. From a public health perspective, we are especially concerned with the lower tail of the birthweight distribution, and in particular with cases that fall below the "low birthweight" threshold of 2500 grams. Koenker and Hallock (2001) show that providing pre-natal care has a much larger impact at lower conditional quantiles than at higher conditional quantiles of the birthweight distribution. This suggests that such practices may provide an effective way of reducing the problem of low birthweight even if they have limited impact on the mean of birthweight.

Starting with Chamberlain (1994) and Buchinsky (1994), quantile regressions have also been used in the large labor economics literature on wage inequality. From a welfare point of view, both the mean and the distribution of real wages matter. It is thus important to see how wage setting factors like education and union status (the two cases studied by Chamberlain) affect both the conditional mean and the conditional dispersion of wages. For instance, Chamberlain shows that unions have a much larger effect at lower than higher quantiles, confirming the well established view that they tend to compress wages among workers with the same observable characteristics.

On closer examination, however, quantile regressions are not always well suited for answering many questions of distributional interest. The key difficulty occurs when the ultimate object of interest is the unconditional distribution, as in the case of low birthweight or wage inequality discussed above. The problem is that a particular quantile of the unconditional outcome distribution cannot be expressed as a function of the corresponding conditional quantiles. Thus the coefficients of a particular quantile regression cannot be used to predict the effect of a given covariate on the corresponding quantile of the unconditional distribution. In fact, we show that even the sign of the coefficient may be "wrong". For example, we show in the empirical section that while unions have a positive effect on the 90th conditional quantile, unionization tends to reduce the 90th quantile of the unconditional wage distribution. By contrast, it is well known that the expected value of the conditional mean is equal to the unconditional mean of the outcome variable. This implies that a regression coefficient can be either interpreted as the effect of a covariate on the conditional or unconditional mean. For instance, if the OLS estimate of the effect of unions on log wages is 0.2, this also means that moving everybody from non-union to union would increase the (unconditional) mean of log wages by 0.2.

Another way of illustrating the limitations of quantile regressions is to go back to the birthweight example. Consider a quantile regression for the 10th conditional quantile. Koenker and Hallock (2001) show that prenatal care has a large effect on this conditional quantile. Unfortunately, the 10th conditional quantile is very different depending on other characteristics of the mother. The 10th quantile for smoking black mothers with a high school education or less (at 2183 grams) is well below the low birthweight threshold of 2500 grams.[1] By contrast, the 10th quantile for white college educated mothers who do not smoke (at 2880 grams) is well above the low birthweight threshold, in fact it is above the corresponding birthweight for the 41th quantile of the previous group. The quantile regression estimate at the 10th conditional quantile thus mixes the impact of prenatal care for some infants above and below the low birthweight threshold. Conditional quantile regressions may not be the best way to infer the impact of prenatal care on the weight of infants right around the 2500 grams threshold.[2]

In this paper, we propose a new computationally simple regression method that can be used to model the unconditional quantiles of the outcome variable as a function of the explanatory variables. Our approach builds upon the concept of the influence function (IF), a widely used tool in robust estimation of statistical or econometric models. Intuitively, the influence function for a given distributional parameter is simply the contribution of an individual observation to this parameter. For example, the influence function of an observation to the mean is simply its demeaned value. More generally, the influence function is easily computed for other distributional measures, including quantiles. Here, we propose, in the simplest case, to run OLS regressions of a (recentered) influence function (RIF) on the explanatory variables. This corresponds to what we usually do in the case of the mean for which the recentered influence function is the observation value itself. The estimated model can thus be used to construct policy conterfactuals or decomposition of the unconditional quantile in the same way OLS is

[1] As computed from the Koenker and Hallock (2001) data.

[2] Empirically however, conditional and unconditional quantile regressions estimates in this particular case will give similar results.

typically used for the mean.

We view our method as a very useful complement to quantile regressions. Of course, in some settings quantile regressions are the appropriate method to use. For instance, quantile regressions are a useful descriptive tool that provide a parsimonious representation of the conditional quantiles. Under the stronger assumption of rank preservation, quantile regression estimates can also be given a causal interpretation as the "treatment effect" for someone of a given rank in the distribution of unobservables. Quantile regressions are also a natural model to estimate in the "traditional" econometric approach where regressors are treated as fixed. We follow a different approach here since we are treating regressors as random, and want to compute the unconditional distribution of the outcome variables under various scenarios about the distribution of regressors.

The structure of the paper is as follows. In the next section, we propose an intuitive introduction to the concept of unconditional quantile regressions by contrasting them with conditional quantile regressions. In Section 3, we establish the general concepts of recentered influence functions which can be applied to any functional of the distribution of interest. We formally show how the recentered influence function can be used to compute what happens to a distributional statistic ν when the distribution of the outcome variable Y changes in response to a change in the distribution of covariates X. In section 4, we focus on the case of quantiles and link the impact of specific changes in the distribution of the covariates on the outcome variables to estimable paramters. We also consider these parameters in the context of structural models. We discuss estimation issues in Section 5. Section 6 illustrates the workings of our method by applying it to the birthweight example above and the impact of unions on the distribution of wages. We conclude in Section 7.

2 Unconditional vs Conditional Quantile Regressions

Before presenting the details of our estimation method, it is important to first clarify the difference between commonly estimated conditional quantile regressions and our proposed approach. As we note above, the key difference is that the parameters from a standard quantile regression indicate the effect of a covariate on a conditional quantile, but in many cases the effect on the unconditional quantile may be of more policy or economic interest. A useful way of illustrating the difference between the conditional and unconditional effects is to go back to the case of the mean for which the literature is extensively developed.

When the conditional expectation of the outcome variable, Y, is linear in the explanatory variables, X, it is straighforward to estimate the partial effect of each element of X by OLS. As is well know, OLS provides consistent estimates of the average partial effect (APE) of X on Y, where the "average" reflects the fact that the effect is averaged over all possible values of Y given X. The simple link between model parameters and APE does not extend, however, to non-linear models. For example, in a probit model, the parameters of the latent model do not correspond to the APE of X on the means of the outcome variable Y (a proportion in this case). As a result, applied researchers commonly compute some average partial (or marginal) effects that are of more direct economic interest than the parameters per se. In some case, people simply compute the partial effect at a particular value of the explanatory variables (for instance the mean of X). Since this conditions on a particular value of X, this particular effect corresponds to what Wooldridge (2004) calls a conditional average partial effect ($CAPE$). In other case, the partial effect is averaged out over all observed values of the explanatory variables, which now corresponds to what Wooldridge (2004) calls an unconditional average partial effect ($UAPE$).

We can similarly think of quantile regressions as a method for estimating quantile partial effects. If those quantile partial effects refer to the conditional quantiles of Y given a particular value X, conditional quantile regressions are typically involved. We refer to this type of quantile partial effects as "conditional quantile partial effects" ($CQPE$). On the other hand, the method proposed here seeks to estimate in a simple fashion the effect of changes in X on unconditional quantiles of Y. We will refer to these partial effects as unconditional quantile partial effects ($UQPE$).

The difference between conditional and unconditional partial effects is easily illustrated in the case of the mean, $\mu = E[Y] = \int y \cdot dF_Y(y)$. From the law of iterated expectations, it follows that

$$\mu = \int E[Y|X = x] \cdot dF_X(x).$$

This shows that the unconditional mean of Y can be recovered by integrating the conditional mean $E[Y|X]$, using the distribution of X. Generally speaking, we are interested in the effect of X on the outcome variable Y. Since $E[Y|X]$ is easily estimated by regression methods, a first answerable question is what is the effect of a small change in X (holding everything else constant) on the conditional expectation $E[Y|X]$ evaluated at a specific value x of X. As mentioned above, Wooldridge calls this effect the conditional average partial effect ($CAPE$) to highlight the fact that it is being computed for a given value of X

$$CAPE(x) = \frac{\partial E\left[Y|X=x\right]}{\partial x}.$$

By contrast, the unconditional APE ($UAPE$) captures the average impact of a small change of X (at all values of X) on the unconditional expectation of Y , $E(Y) = \mu$. If follows from the law of iterated expectations that

$$UAPE = \int \frac{\partial E\left[Y|X=x\right]}{\partial x} \cdot dF_X(x) = \int CAPE(x) \cdot dF_X(x).$$

The last part of the equation shows that the $UAPE$ also turns out be the average of $CAPE(x)$ over all values of X. As we show later, however, this interesting result does not generalize beyond the mean.

In the standard linear regression model where the conditional expectation is assumed to be linear in X [$E(Y|X) = X'\beta$], the $CAPE$ is equal to β for all values of X, so that $UAPE = CAPE$. This results has great practical importance. It means that the OLS estimate of β always provides the "correct" answer, irrespective of whether we are interested in estimating the conditional *or* the unconditional effect of X on the mean of Y. This result does not hold, however, for other functionals of Y and, in particular, for quantiles. When the focus is on quantiles, which are the object of interest in many studies about distributional issues, one has to be much more careful in deciding whether the question of economic or policy interest has to do with conditional or unconditional effects.

Since quantile regressions are a model of the conditional distribution of Y given X, they provide direct estimates of the $CQPE$. Unlike the case of the mean, however, the integration of those conditional effects using the distribution of X does not yield any meaningful functional of the unconditional distribution of Y. By contrast, our proposed regression method provides direct estimates of the partial effect of X on the unconditional quantile of Y, the so-called $UQPE$. For each possible value of X, the conditional τ^{th} quantile of Y, $Q_\tau(Y|X)$, is implicitly defined as

$$\tau = \left[\mathbb{I}\left\{y \leq Q_\tau(Y|X)\right\}|X\right],$$

while the unconditional τ^{th} quantile, q_τ, is defined as

$$\tau = \int E\left[\mathbb{1}\left\{Y \leq q_\tau\right\}|X\right] \cdot dF_X(x).$$

Unlike the case of the mean, however, there is no "law of iterated quantiles" that can provide a relationship between the conditional and unconditional quantiles. This means that, in general

$$q_\tau \neq \int Q_\tau(Y|x) \cdot dF_X(x).$$

Taking derivatives from both sides of the equation, it follows that the integration of the conditional quantile partial effects with respect to the distribution of X bears no relationship to the parameter investigated in this article, the "unconditional quantile partial effects" ($UQPE$):

$$UQPE \neq \int CQPE\left(x\right) \cdot dF_X(x) = \int \frac{\partial Q_\tau(Y|x)}{\partial x} \cdot dF_X(x)$$

To define the $UQPE$ properly, consider the general case of a distributional statistic ν, such as the mean or quantiles. As we later show, it is always the case that:

$$\nu = \int \mathrm{RIF}(y;\nu) \cdot dF_Y(y) = \int E\left[\mathrm{RIF}(Y;\nu)|X = x\right] \cdot dF_X(x)$$

where $\mathrm{RIF}(Y;\nu)$ is the *recentered influence function*, which we formally defined in the next section. The $\mathrm{RIF}(Y;\nu)$, besides having an expected value equal to functional ν, corresponds to the leading term of a von Mises type expansion. The parameter $UQPE$ is simply the expected value of the derivative with respect to x (the average derivative) of the conditional expectation of $\mathrm{RIF}(Y;q_\tau)$ given $X = x$, that is,

$$UQPE = \int \frac{\partial E\left[\mathrm{RIF}(Y;q_\tau)|X = x\right]}{\partial x} \cdot dF_X(x)$$

The conditional expectation $E\left[\mathrm{RIF}(Y;q_\tau)|X\right]$ can be estimated by the same regression methods used in the case of the mean. In particular, when RIF is assumed to be linear in X, ($E\left[\mathrm{RIF}(Y;q_\tau)|X\right] = X'\gamma_\tau$), then the $UQPE$ is simply the regression parameter γ_τ of $\mathrm{RIF}(Y;q_\tau)$ on the X.[3] Generally speaking, the $UQPE$ can be thought of as a specific

[3]In case of the mean, even if $E\left[\mathrm{RIF}(Y;\mu)|X = x\right]$ is non-linear in x, the UAPE has both the interpretations of being the average of conditional effects and the partial effect on the unconditional average. This results from the fact that $\mathrm{RIF}(Y;\mu) = Y$.

policy experiment where X is increased by an equal small amount for all observations. There is no particular reason, however, why it should always be the parameter of interest. Our method can easily address other policy experiments.

To better illustrate the workings of the method for quantiles, which are the focus of this paper, we now give an intuitive interpretation of the recentered influence function, $\mathrm{RIF}(Y; q_\tau)$, for the τ^{th} quantile, q_τ:

$$\mathrm{RIF}(Y; q_\tau) = q_\tau + \mathrm{IF}(Y; q_\tau) = q_\tau + \frac{\tau - \mathbb{1}\{Y \leq q_\tau\}}{f_Y(q_\tau)}. \tag{1}$$

The derivation and theoretical foundations of this expression are shown below. First note that the recentered influence function $\mathrm{RIF}(Y; q_\tau)$ only depends on the value of Y through the indicator variable $\mathbb{1}\{Y \leq q_\tau\}$. Thus, observations with a value of Y below q_τ have a negative influence on q_τ, while observations with a value of Y above q_τ have a positive influence on q_τ. Leaving aside the constant τ and the density (scaling factor) $f_Y(q_\tau)$, this implies that a regression of the influence function on covariates is simply a linear probability model for whether a given observation lies above or below the quantile q_τ. Alternatively, one can run a logit (or probit) model and compute the average marginal effects of each covariate. Since the indicator variable is divided by the density $f_Y(q_\tau)$ in equation (1), the coefficients in the influence function regression are simply the coefficients of the linear probability model rescaled by the density $f_Y(q_\tau)$, which can be easily estimated using kernel density methods.

The intuition for why we need to divide the linear probability model coefficients by the density is easily understood in the context of the birthweight example.[4] Say we run a linear probability model for whether birthweight is below the low birthweight threshold of 2500 grams. The estimated coefficient on prenatal care now indicates by how much prenatal care reduces the incidence of low birthweight. Say that it reduces the incidence of low birthweight by one percentage point, from 6 to 5 percent. While this probability impact is of some clear interest, it is not directly comparable to the corresponding estimated effect of prenatal care in an OLS or quantile regression of birthweight. To get a comparable effect, we need to transform the probability impact back into a birthweight impact. Roughly speaking, the function that transforms probabilities into birthweights is simply the inverse of the cumulative distribution function (CDF), and the slope of the inverse CDF is the inverse of the density. If the density of birthweight at 2500 grams is

[4] We note that the methods for binary choice model proposed by Lewbel (2004, 2005) also involve weighting by the inverse probability density function.

1/10,000, then the 0.01 impact of prenatal care on the probability of low birthweight will translate into a 0.01×10,000=100 grams impact on the 10^{th} quantile of the birthweight distribution. Prenatal care would thus increase the 10^{th} quantile of the birthweight distribution from 2500 to 2600 grams. The reader primarily interested in applications can skip to Section 6 to get a precise idea of what empirical $UQPE$ and $CQPE$ look like.

3 General Concepts

In this section we first review the concept of influence functions, which arises in the von Mises (1947) approximation and is largely used in the robust statistics literature. We then introduce the concept of *recentered influence functions* which will be key to the derivation of unconditional quantile regressions. Finally we apply the von Mises approximation, defined for a general alternative or counterfactual distribution, to the case of where this counterfactual distribution arises from changes in the covariates. Using the law of iterated expectations, we can write the derivative of ν in the direction of the counterfactual distribution of interest as the integral of the conditional expectation of the recentered influence function given X, over the distributional change of interest in the covariates. The derivations are developed for general functionals of the distribution; they will be applied to quantiles and the mean, for comparison, in the next section.

3.1 Definition of Recentered Influence Functions

We begin by recalling the theoretical foundation of the definition of the influence functions following Hampel et all (1986). Hampel (1968, 1974) introduced the influence function as a measure to study the infinitesimal behavior of real-valued functionals $\nu(F)$, where $\nu : \mathcal{F}_\nu \to \mathbb{R}$ and $\mathcal{F}_\nu$ is a class of distribution functions such that $F \in \mathcal{F}_\nu$ if $|\nu(F)| < +\infty$. Following Huber (1977), we say that $\nu(\cdot)$ is Gâteaux diferentiable at F if there exists a real kernel function $a(\cdot)$ such that for all G in $\mathcal{F}_\nu$:

$$\lim_{t\downarrow 0} \frac{\nu(F_{t,G}) - \nu(F)}{t} = \frac{\partial \nu(F_{t,G})}{\partial t}|_{t=0} = \int a(y) \cdot dG(y) \tag{2}$$

where $0 \leq t \leq 1$ and where the mixing distribution $F_{t,G}$

$$F_{t,G} = (1-t) \cdot F + t \cdot G = t \cdot (G - F) + F \tag{3}$$

is the probability distribution that is t away from F in the direction of the probability distribution G.

The expression on the left hand side of equation (2) is the directional derivative of ν at F in the direction of G. Note that we can replace $dG(y)$ on the right hand side of equation (2) by $d(G-F)(y)$:

$$\lim_{t\downarrow 0}\frac{\nu\left((1-t)\cdot F+t\cdot G\right)-\nu(F)}{t}=\frac{\partial\nu\left(F_{t,G}\right)}{\partial t}|_{t=0}=\int a(y)\cdot d(G-F)(y) \tag{4}$$

since $\int a(y)\cdot dF(y)=0$, which can be shown by considering the case where $G=F$.

The concept of influence function arises from the special case in which G is replaced by Δ_y , the probability measure that put mass 1 at the value y, in the mixture $F_{t,G}$. This yields F_{t,Δ_y}, the distribution that contains a blip or a contaminant at the point y,

$$F_{t,\Delta_y}\equiv(1-t)\cdot F+t\cdot\Delta_y$$

The influence function of the functional ν at F for a given point y is defined as

$$\begin{aligned}\mathrm{IF}(y;\nu,F) &\equiv \lim_{t\downarrow 0}\frac{\nu(F_{t,\Delta_y})-\nu(F)}{t}=\frac{\partial\nu\left(F_{t,\Delta_y}\right)}{\partial t}|_{t=0}\\ &= \int a(y)\cdot d\Delta_y(y)=\ a(y)\end{aligned} \tag{5}$$

Combining (4) and (5) explicitly shows that the directional derivative of the functional $\nu\left(F_{t,G}\right)$, at F with a small contaminant in the direction of the distribution G, is obtained by integrating up the influence function at F over the distributional differences between G and F, which arise from the introduction of the contaminant,

$$\frac{\partial\nu\left(F_{t,G}\right)}{\partial t}|_{t=0}=\int\mathrm{IF}(y;\nu,F)\cdot d(G-F)(y)\,. \tag{6}$$

Using the definition of the influence function, the functional $\nu\left(F_{t,G}\right)$ itself can be represented as a von Mises linear approximation (VOM):[5]

$$\nu\left(F_{t,G}\right)=\nu(F)+t\cdot\int\mathrm{IF}(y;\nu,F)\cdot d(G-F)(y)+r(t;\nu;G,F) \tag{7}$$

[5]This expansion can be seen as a Taylor series approximation of the real function $A(t)=\nu\left(F_{t,G}\right)$ around $t=0:\ A(t)=A(0)+A'(0)\cdot t+Rem_1$. But since $A(0)=\nu(G)$, and $A'(0)=\int a_1(y)\,d(G-F)(y)$, where $a_1(y)$ is the influence function, we get the VOM approximation.

where $r(t;\nu;G,F)$ is a remainder term.[6] Now consider the leading term of equation (7) as an approximation for $\nu(G)$, that is, for $t=1$:

$$\nu(G) \approx \nu(F) + \int \mathrm{IF}(y;\nu,F) \cdot dG(y).$$

By analogy with the influence function, for the particular case $G=\Delta_y$, we call this first order approximation term the *Recentered Influence Function* (RIF)

$$\mathrm{RIF}(y;\nu) = \nu(F) + \int \mathrm{IF}(y;\nu,F) \cdot d\Delta_y(y) = \nu(F) + \mathrm{IF}(y;\nu,F).$$

The recentered influence function $\mathrm{RIF}(y;\nu)$ has two interesting properties: i) it integrates up to the functional $\nu(F)$ and ii) provides a simple way to obtain the asymptotic variance of the functional $\nu(F)$ (as does the influence function)

$$\begin{array}{ll} i) & \int \mathrm{RIF}(y;\nu) \cdot dF(y) = \int (\nu(F) + \mathrm{IF}(y;\nu,F)) \cdot dF(y) = \nu(F) \\ ii) & \int (\mathrm{RIF}(y;\nu) - \nu(F))^2 \cdot dF(y) = \int (\mathrm{IF}(y;\nu,F))^2 \cdot dF(y) = AV(\nu,F) \end{array}$$

where $AV(\nu,F)$ is the asymptotic variance of functional ν under the probability distribution F.

Note also that the equivalent of equation (6) holds[7]

$$\frac{\partial \nu(F_{t,G})}{\partial t}|_{t=0} = \int \mathrm{RIF}(y;\nu,F) \cdot d(G-F)(y).$$

This result would not be very useful if both F and G were known, since we could then compare $\nu(G)$ and $\nu(F)$ to see how the distribution parameter ν changes when the distribution moves from F to G. What we have in mind here, however, is that while F is the observed distribution of Y, G is an alternative counterfactual distribution.

[6] See Withers (1983), Fernholz (2001) for the study of higher order terms. The second order term is much more involved

$$\begin{aligned} A''(0) &= \int\int a_2(x,y)\, d(G-F)(x) d(G-F)(y) \\ \text{where} \quad a_2(x,y) &= \frac{\partial}{\partial t}\mathrm{IF}(y;\nu,F_{t,\Delta y})|_{t=0} + \mathrm{IF}(y;\nu,F)), \end{aligned}$$

and where $a_2(x,y) = a_2(y,x)$, for all x,y.

[7] This follows from the fact that densities integrate to one.

3.2 Impact of General Changes in the Distribution of X

The counterfactual distribution G of interest here will be the one that prevails when we change the distribution of covariates X. Specifically, assume that we observe Y in the presence of covariates X, so that Y and X have a joint distribution, $F_{Y,X}(\cdot,\cdot):\mathbb{R}\times\mathcal{X}\rightarrow[0,1]$, and $\mathcal{X}\subset\mathbb{R}^k$ is the support of X. We can now talk of the impact of distributional changes in the covariates X on Y, since any change a covariate X can be represented as a change in the distribution of the X. The recentered influence function will provide a convenient way of assessing the impact of changes in the covariates on the distribution statistic ν without having to compute the corresponding counterfactual distribution which is, in general, a difficult estimation problem.

We begin by considering general changes in the covarates, represented by $G_X(x)$ the counterfactual distribution of the covariates X of desired changes. For example, we want to know how $\nu(F_Y)$, the functional of the unconditional distribution of Y, is affected if we allow a change of size t in the distribution of covariates X from F_X in the direction of G_X, where G_X is another probability distribution defined over $\mathcal{X}$ and t is a positive real number, which can be made arbitrarily small. We will narrow our attention to specific distributional changes in the covariates and their associated impacts on Y in Section 4.

These general distributional changes can be described by mixtures of distributions that we now define formally.[8] By definition, the unconditional distribution function of Y can be written as:

$$F_Y(y)=\int F_{Y|X}(y|X=x)\cdot dF_X(x)$$

where the subscript Y is now used to differentiate the unconditional distribution, F_Y, from the conditional distribution, $F_{Y|X}$, and the distribution of covariates, F_X. Thus it is always the case, by the law of iterated expectations, that

$$\nu(F_Y)=\int \mathrm{RIF}(y;\nu)\cdot dF_Y(y)=\int E\left[\mathrm{RIF}(Y;\nu)|X=x\right]\cdot dF_X(x)$$

where $E\left[\mathrm{RIF}(Y;\nu)|X=x\right]=\int_y \mathrm{RIF}(y;\nu)\cdot dF_{Y|X}(y|X=x)$.

The counterfactual mixing distribution $F_{Y,t\cdot G_X}(y)$ represent the distribution of y resulting from a change t away from F_X in the direction of the alternative probability distribution G_X

$$F_{Y,t\cdot G_X}(y)\equiv\int F_{Y|X}(y|X=x)\cdot dF_{X,t\cdot G_X}(x)$$

[8]Note that the following integral operators correspond to k integrals, as X is now assumed to be a random vector of length k.

where $F_{X,t\cdot G_X}$ is the mixing distribution that captures a change t away from F_X in the direction of the probability distribution G_X.[9] The counterfactual distribution G_Y is obtained similarly by replacing the distribution of covariates, F_X, by the alternative distribution, G_X:

$$G_Y(y) \equiv \int F_{Y|X}(y|X=x) \cdot dG_X(x)$$

that is, the distribution G_Y now stands explicitly for the counterfactual distribution, as opposed to the general distribution G used above.

Our central result here is that the impact of a small change of size t in the distribution of covariates from F_X in the direction of G_X on the functional $\nu(F_Y)$ is given by integrating up the expectation of the recentered influence function with respect to the changes in distribution of the covariates $d(G_X - F_X)$.

THEOREM 1 [***Effect of Infinitesimal Distributional Changes*** $t \cdot G_X$]***:***

$$\begin{aligned} \frac{\partial \nu(F_{Y,t\cdot G_X})}{\partial t}|_{t=0} &= \int \mathrm{RIF}(y;\nu) \cdot d(G_Y - F_Y)(y) \\ &= \int E[\mathrm{RIF}(Y;\nu)|X=x] \cdot d(G_X - F_X)(x) \end{aligned}$$

The proof, provided in the Appendix A, builds on the fact that since, by definition, G_Y only changes in response to a change in G_X, the mixing distribution $F_{Y,t\cdot G_Y}(y) = F_{Y,t\cdot G_X}(y)$. Because the result applies to any distribution of interest G_X, this theorem gives us a powerful tool to consider commonly estimated partial effects, as well as effects related to discrete changes linked to any policy of interest. The conditional expectation $E[\mathrm{RIF}(Y;\nu)|X]$ can be estimated by the same regression methods used in the case of the mean. In particular, we show in Section 5 that, in the case of quantiles, if RIF is assumed to be linear in X, $(E[\mathrm{RIF}(Y;q_\tau)|X] = X'\gamma_\tau)$, then the $UQPE$ will simply the regression parameter γ_τ of $\mathrm{RIF}(Y;q_\tau)$ on the X. We also show that, alternatively, non-parametric estimation methods can be applied.

4 Application to Unconditional Quantiles

We now turn to the application of these general results to the case of unconditional quantiles, while providing a comparison with the case of the mean. We will also consider two important specific cases of the impact of changes in the covariates. The leading case

[9] Or more formally, $F_{X,t\cdot G} \equiv (1-t) \cdot F_X + t \cdot G_X$.

of such a specific change is the unconditional quantile partial effect $UQPE$ defined in Section 2, the second case $UQPE(x)$ correspond to an unconditional quantile partial effect evaluated at a specific value of $X = x$. We also show that for some structural models, there is a direct link between the $UQPE$ and the structural parameters.

4.1 Recentered Influence Functions for Quantiles

In this section, we explicitly derive the recentered influence function for quantiles and show that it provides an approximation of order (t) to the actual quantile. However to fix ideas, we begin with the familiar case of the mean, $\nu(F) = \mu$, for which the concept of the recentered influence function is trivial and where the approximation in equation (7) is exact.

Applying the definition of influence function of equation (5) to $\mu = \int y \cdot dF(y)$, we find that the influence function of the mean at a point y is simply that point itself, but demeaned

$$\begin{aligned}
\mathrm{IF}(y;\mu,F) &= \frac{\partial \nu\left(F_{t,\Delta_y}\right)}{\partial t}|_{t=0} = \frac{\partial \int y \cdot ((1-t)\cdot dF(y) + t \cdot d\Delta_y(y))}{\partial t}|_{t=0} \\
&= \frac{\partial \left(t \cdot \int y \cdot d(\Delta_y - F)(y) + \int y \cdot dF(y)\right)}{\partial t}|_{t=0} \\
&= \int y \cdot d(\Delta_y - F)(y) = y - \int y \cdot dF(y) \\
&= y - \mu
\end{aligned}$$

We also find when we apply the VOM linear approximation of equation (7) to the mean, μ, that the remainder $r(t;\mu;G,F)$ equals zero

$$\begin{aligned}
\nu(G) &= \nu(F) + \int \mathrm{IF}(y;\nu,F)\cdot d(G-F)(y) + r(t;\nu;G,F) \\
&= \mu + \int (y-\mu)\cdot dG(y) + r(t;\mu;G,F) \\
&= \int y \cdot dG(y) + r(t;\mu;G,F) = \nu(G) + r(t;\mu;G,F).
\end{aligned}$$

The recentered influence function of the mean is thus trivial

$$\mathrm{RIF}(y;\mu) = \mu + \mathrm{IF}(y;\mu,F) = \mu + y - \mu = y$$

so are the expressions for the expectation of the $\mathrm{RIF}(y;\mu)$ and its asymptotic variance :

$$\begin{array}{ll} i) & \int \mathrm{RIF}(y;\mu)\cdot dF(y) = \int y\cdot dF(y) = \mu \\ ii) & \int \left(\mathrm{RIF}(y;\mu)-\mu\right)^2\cdot dF(y) = \int (y-\mu)^2\cdot dF(y) = \sigma^2 \end{array}$$

Turning to our application of interest, the expression for the influence function of quantiles also has some intuitive appeal: it is equal to a function that locates the point y either above or below quantile τ, inversely weighted by the density at that point. Consider the τ^{th} quantile: $\nu(F)=q_\tau$, which is defined implicitly as the integral bound in

$$\tau = \int_{-\infty}^{q_\tau} dF(y) = \int_{-\infty}^{\nu(F)} dF(y) = \int_{-\infty}^{\nu(F_{t,\Delta_y})} dF_{t,\Delta_y}(y).$$

It is easily shown by taking the derivative of this last expression with respect to t and rewriting the resulting expression in terms of $\partial\nu\left(F_{t,\Delta_y}\right)/\partial t$ that[10]

$$\begin{aligned} \mathrm{IF}(y;q_\tau,F) &= \frac{\partial\nu\left(F_{t,\Delta_y}\right)}{\partial t}|_{t=0} \\ &= \frac{\int \mathbb{I}\{y\le\nu(F)\}\,dF(y) - \mathbb{I}\{y\le\nu(F)\}}{dF(y)/dy|_{y=\nu(F)}} \\ &= \frac{\tau-\mathbb{I}\{y\le q_\tau\}}{f(q_\tau)}, \end{aligned}$$

where $f(\cdot)$, is the probability density associated with the probability distribution F.

In the case of the quantile, the VOM appproximation does not hold exactly:

$$\begin{aligned} \nu(G) &= \nu(F) + \int \mathrm{IF}(y;\nu,F)\cdot d(G-F)(y) + r(t;\nu;G,F) \\ &= q_\tau + \int \left(\frac{\tau-\mathbb{I}\{y\le q_\tau\}}{f(q_\tau)}\right)\cdot dG(y) + R(t;q_\tau;G,F) \end{aligned}$$

rather the reminder is of the order t

$$\begin{aligned} R(t;q_\tau;G,F) &= t\cdot\left(\int\left(\frac{\tau-\mathbb{I}\{y\le\widetilde{q}_{\tau,G}\}}{f_G(\widetilde{q}_{\tau,G})}\right)\cdot dF(y) - \int\left(\frac{\tau-\mathbb{I}\{y\le q_\tau\}}{f(q_\tau)}\right)\cdot dG(y)\right) \\ &\quad + r(t;\nu;G,F) \\ &= O(t) \end{aligned}$$

[10]See the Appendix A for details

since $r(t;\nu;G,F)$ is $o(t)$, where $\widetilde{q}_{\tau,G}=\nu(G)$ when ν is the τ^{th}quantile and $f_G(\cdot)$ is the probability density function associated with the probability distribution G.

The recentered influence function is

$$\text{RIF}(y;q_\tau)=q_\tau+\text{IF}(y;q_\tau,F)=q_\tau+\frac{\tau-\mathbb{1}\{y\leq q_\tau\}}{f(q_\tau)}.$$

It is easy to check that the mean of the recentered influence function of q_τ is the quantile itself:

$$\int \text{RIF}(y;q_\tau)\cdot dF(y)=\int\left(q_\tau+\frac{\tau-\mathbb{1}\{y\leq q_\tau\}}{f(q_\tau)}\right)\cdot dF(y)=q_\tau$$

and its variance is:

$$\int(\text{RIF}(y;q_\tau)-q_\tau)^2\cdot dF(y)=\int\left(\frac{\tau-\mathbb{1}\{y\leq q_\tau\}}{f(q_\tau)}\right)^2\cdot dF(y)=\frac{\tau\cdot(1-\tau)}{f^2(q_\tau)}$$

4.2 Impact of Changes in the Distribution of X on Unconditional Quantiles

Let us start by applying the general definitions of section 3.2 to the case of quantiles. When X is multivariate and continuously distributed, we can express the quantile τ in terms of the conditional distribution of Y given X and of the distribution of X

$$\tau=F_Y(q_\tau)=\int F_{Y|X}(q_\tau|X=x)\cdot dF_X(x).$$

Then the counterfactual quantile $\widetilde{q}_{\tau,t\cdot G_X}(t)$ that would result from the small change t in the direction of G_X, is implicitly defined by

$$\tau=\int^{\widetilde{q}_{\tau,t\cdot G_Y}(t)} dF_{Y,t\cdot G_X}(y)=\int F_{Y|X}\left(\widetilde{q}_{\tau,G_X}(t)\,|X=x\right)\cdot dF_{X,t\cdot G_X}(x)$$

and the impact of the infinitesimal change $t\cdot G_X$ itself simply follows from the application of Theorem 1:

$$\begin{aligned}\frac{\partial\widetilde{q}_{\tau,t\cdot G_X}(t)}{\partial t}|_{t=0} &= \int E\left[\text{RIF}(Y;q_\tau)|X=x\right]\cdot d\left(G_X-F_X\right)(x)\\ &= -\frac{1}{f_Y(q_\tau)}\cdot\int\left(F_{Y|X}(q_\tau|X=x)\right)\cdot d\left(G_X-F_X\right)(x)\end{aligned}$$

4.2.1 Unconditional Quantile Partial Effects

Let's now focus of some specific changes of interest. We show that the Unconditional Quantile Partial Effects (UQPE) correspond to the policy experiment where everybody in the population is given an infinitesimal ξ increase in their X values, where $\xi \in \mathbb{R}^k$. For this experiment to be meaningful, assume that support of X is unbounded, or at least, not bounded above. Thus, the distribution G_X will be that of the random variable $Z = X + \xi$. The cumulative distribution function of Z is simply $F_Z(x) = F_X(x - \xi)$.[11]

To characterize the effects of changing the distribution of X from F_X to F_Z on q_τ, the τ-th quantile of the unconditional distribution of Y, consider first the marginal effect on the unconditional quantile of increasing X by ξ and call this parameter $\kappa(q_\tau; \xi)$

$$\begin{aligned}\kappa(q_\tau; \xi) &= \frac{\partial \widetilde{q}_{\tau, t \cdot F_X(x-\xi)}(t)}{\partial t}|_{t=0} = \int E[\text{RIF}(Y; q_\tau)|X = x] \cdot (dF_X(x - \xi) - dF_X(x)) \\ &= \int (E[\text{RIF}(Y; q_\tau)|X = x + \xi] - E[\text{RIF}(Y; q_\tau)|X = x]) \cdot dF_X(x)\end{aligned}$$

then letting ξ go to zero

$$UQPE(\tau) = \frac{\partial \kappa(q_\tau; \xi)}{\partial \xi}|_{\xi=0} = \int \frac{dE[\text{RIF}(Y; q_\tau)|X = x]}{dx} \cdot dF_X(x) \tag{8}$$

$$= -\frac{1}{f_Y(q_\tau)} \cdot \int F'_{Y|X}(q_\tau|X = x) \cdot dF_X(x) \tag{9}$$

where $F'_{Y|X}(q_\tau|X = x) = dF_{Y|X}(q_\tau|X = x)/dx$ is the vector of partial derivatives, in which each entry corresponds to $\partial F_{Y|X}(q_\tau|X = x)/\partial x_j,\ j = 1, \ldots, J$. In the case of discrete covariates where X may assume K different values, we define $UQPE(\tau)$ as[12]

$$\begin{aligned}UQPE(\tau) &= \sum_{k=1}^{K} \left(\frac{E[\text{RIF}(Y; q_\tau)|X = x_{k+1}] - E[\text{RIF}(Y; q_\tau)|X = x_k]}{x_{k+1} - x_k} \right) \cdot \Pr[x_k] \\ &= -\frac{1}{f_Y(q_\tau)} \cdot \sum_{k=1}^{K} \left(\frac{F_{Y|X}(q_\tau|X = x_{k+1}) - F_{Y|X}(q_\tau|X = x_k)}{x_{k+1} - x_k} \right) \cdot \Pr[x_k]\end{aligned}$$

While $\kappa(q_\tau; \xi)$ corresponds exactly the marginal effect on q_τ of changing the distribution of X from F_X to F_Z, it may not be very interesting as it depends on ξ, the value

[11]Note the important particular case in which $\xi = [0, \ldots, 0, 1, 0, \ldots, 0]'$ corresponds to an unit increase in the covariate X_j.

[12]In what follows, by a normalization argument we make $x_{K+1} = x_K$ and therefore $F_{Y|X}(q_\tau|X = x_{K+1}) = F_{Y|X}(q_\tau|X = x_K)$ and $E[\text{RIF}(Y; q_\tau)|X = x_{K+1}] = E[\text{RIF}(Y; q_\tau)|X = x_K]$.

of the transfer. Its infinitesimal version $UQPE(\tau)$ is the average partial effect on q_τ of a small lump-sum changes in the distribution of X. As we show below, it corresponds in a number of settings to the usual parameter of interest.

4.2.2 Unconditional Quantile Partial Effects at $X = x$

There are some circumstances, such as the effect of compulsory schooling on the distribution of earnings for example, where the policy maker is particularly interested in the effect of some fixed value x of the covariate X on the quantile τ of the unconditional distribution of Y. To capture such effect, for X continuously distributed, we need to compare the effects of the two following experiments on the quantile of the marginal distribution of Y: a change towards $X = x + \xi$ and a change towards $X = x$. These two experiments correspond to changes in the distribution of X from F_X to Δ_x and $\Delta_{x+\xi}$ the degenerate distributions at x and $x + \xi$, respectively. Analogously to the $UQPE$ parameter, we define the following parameter that allow the comparison between these two experiments:

$$\kappa(q_\tau; x, \xi) = \frac{\partial \widetilde{q}_{\tau, t\cdot\Delta_{x+\xi}}(t)}{\partial t}|_{t=0} - \frac{\partial \widetilde{q}_{\tau, t\cdot\Delta_x}(t)}{\partial t}|_{t=0}$$

$$\begin{aligned} \text{and} \quad UQPE(\tau; x) &= \frac{\partial \kappa(q_\tau; x, \xi)}{\partial \xi}|_{\xi=0} = \int \frac{dE[\mathrm{RIF}(Y; q_\tau)|X = x]}{dx} \cdot d\Delta_x(x) \\ &= \frac{dE[\mathrm{RIF}(Y; q_\tau)|X = x]}{dx} = -\frac{1}{f_Y(q_\tau)} \cdot F'_{Y|X}(q_\tau|X = x). \end{aligned}$$

In the case of discrete covariates where X may assume K different values, we define $UQPE(\tau; x_k)$

$$\begin{aligned} UQPE(\tau; x_p) &= \frac{E[\mathrm{RIF}(Y; q_\tau)|X = x_{k+1}] - E[\mathrm{RIF}(Y; q_\tau)|X = x_k]}{x_{k+1} - x_k} \\ &= -\frac{1}{f_Y(q_\tau)} \cdot \frac{F_{Y|X}(q_\tau|X = x_{k+1}) - F_{Y|X}(q_\tau|X = x_k)}{x_{k+1} - x_k}. \end{aligned}$$

The policy parameter $UQPE(\tau; x)$, is the unconditional quantile partial effect evaluated at a given $X = x$.[13] It is a second derivative and corresponds exactly to the

[13] Note that this parameter is different from $CQPE$ introduced in section 2 which is based on the conditional quantile rather $UQPE(\tau; x)$ is based on the unconditional quantile of Y.

integrand in the expression derived for $UQPE(\tau)$, that is, for the continuous case

$$UQPE(\tau) = \int UQPE(\tau; x) \cdot dF_X(x)$$

or in general

$$UQPE(\tau) = E[UQPE(\tau; X)].$$

It will prove a useful tool to understand the interpretations of these parameters under various structural assumptions below.

4.2.3 A brief comparison with the case of the mean μ

We now verify that the formulas above for the effects of changes in the distribution of X when applied to μ the unconditional expectation of Y, for which $\text{RIF}(Y;\mu) = Y$, give us the well-known parameters, $UAPE$ and $CAPE$. The $UAPE$ is the average derivative of the conditional expectation of Y given X, and $CAPE$, its the conditional counterpart, is simply the derivative of the same conditional expectation:

$$\begin{aligned} UAPE &= \int \frac{dE[\text{RIF}(Y;\mu)|X=x]}{dx} \cdot dF_X(x) \\ &= \int \frac{dE[Y|X=x]}{dx} \cdot dF_X(x) \end{aligned}$$

$$CAPE(x) = dE[\text{RIF}(Y;\mu)|X=x]/dx = dE[Y|X=x]/dx$$

and note finally, as expected, that

$$UAPE = \int CAPE(x) \cdot dF_X(x).$$

4.3 The policy parameter and the structural form

In this subsection, we introduce unobservables variables and assume that the dependent variable Y is a function of observables X and unobservables ε, according to the following model: $g(X, \varepsilon)$ where $g(\cdot,\cdot)$ is an unknown mapping, assumed to be monotonic in ε. We will consider two situations: independence between X and ε; and no independence. We then link the parameters $UQPE(\tau)$ and $UQPE(\tau; x)$ to this structural model.

We first derive a general expression that links the partial effect of changes in X_j on the marginal quantile of Y to the structural form $Y = g(X, \varepsilon)$ through the $UQPE(\tau; x)$ parameter, as the $UQPE(\tau)$ follows by integrating the distribution of covariates.

Proposition 1 [***UQPE and its relation to the structural form***]:
i) Assuming that the structural form $Y = g(X, \varepsilon)$ is monotonic in ε, , the function $UQPE_j(\tau; x)$, which corresponds to the j^{th} entry in the J-vector $dE[\mathrm{RIF}(Y; q_\tau)|X = x]/dx$ of partial derivatives can be expressed as:

$$\begin{aligned} UQPE_j(\tau; x) &= \frac{\partial E[\mathrm{RIF}(Y; q_\tau)|X = x]}{\partial x_j} \\ &= \left[\frac{\partial g(x, q_\tau)}{\partial x_j} \cdot \frac{f_{\varepsilon|X}(g^{-1}(x, q_\tau))}{\partial g(x, q_\tau)/\partial q_\tau} - \frac{\partial F_{\varepsilon|X}(\cdot|x)}{\partial x_j}\right] \Big/ \int \frac{f_{\varepsilon|X}(g^{-1}(x, q_\tau))}{\partial g(x, q_\tau)/\partial q_\tau} \cdot dF_X(x) \end{aligned}$$

while its expectation, the parameter $UQPE_j(\tau)$, will be:

$$\begin{aligned} UQPE_j(\tau) &= E\left[\frac{\partial g(X, q_\tau)}{\partial x_j} \cdot \frac{f_{\varepsilon|X}(g^{-1}(X, q_\tau))}{\partial g(X, q_\tau)/\partial q_\tau} - \frac{\partial F_{\varepsilon|X}(g^{-1}(X, q_\tau)|X)}{\partial x_j}\right] \\ &\quad \cdot E\left[\frac{f_{\varepsilon|X}(g^{-1}(X, q_\tau))}{\partial g(X, q_\tau)/\partial q_\tau}\right]^{-1} \\ &= E\left[\frac{\partial g(X, q_\tau)}{\partial x_j} \cdot \frac{f_\varepsilon(g^{-1}(X, q_\tau))}{\partial g(X, q_\tau)/\partial q_\tau}\right] \cdot E\left[\frac{f_\varepsilon(g^{-1}(X, q_\tau))}{\partial g(X, q_\tau)/\partial q_\tau}\right]^{-1} \end{aligned}$$

ii) If, in addition, we assume that X and ε are independent, then

$$UQPE_j(\tau; x) = \frac{\partial E[\mathrm{RIF}(Y; q_\tau)|X = x]}{\partial x_j} \quad (10)$$

$$= \frac{\partial g(x, q_\tau)}{\partial x_j} \cdot \frac{f_\varepsilon(g^{-1}(x, q_\tau))}{\partial g(x, q_\tau)/\partial q_\tau} \Big/ \int \frac{f_\varepsilon(g^{-1}(x, q_\tau))}{\partial g(x, q_\tau)/\partial q_\tau} \cdot dF_X(x) \quad (11)$$

while its expectation, the parameter $UQPE_j(\tau)$, will be simply:

$$UQPE_j(\tau) = E\left[\frac{\partial g(X, q_\tau)}{\partial x_j} \cdot \frac{f_\varepsilon(g^{-1}(X, q_\tau))}{\partial g(X, q_\tau)/\partial q_\tau}\right] \cdot E\left[\frac{f_\varepsilon(g^{-1}(X, q_\tau))}{\partial g(X, q_\tau)/\partial q_\tau}\right]^{-1}$$

The proof follows from the application of the definition of the recentered influence function for the quantile and is provided in the Appendix. Under the hypothesis that X and ε are independent and g monotonic in ε, we may invoke the results by Matzkin (2003) that guarantee that both the distribution of ε and the link function g will be non-parametrically identified. Thus, we know that under the independence assumption, our parameters $UQPE_j(\tau; x)$ and $UQPE_j(\tau)$ are identified.

The general formula of Proposition 1 allows us to establish a simple link between

$UQPE$ and the structural parameters of interest in some important special cases that are discussed next.

4.3.1 Case 1: Linear, additively separable model

We start considering that $g(X, \varepsilon) = X^\intercal\beta + \varepsilon$ and that X and ε are independent. Then, $\partial g(X, q_\tau)/\partial X_j = \beta_j$, $\partial g(x, q_\tau)/\partial q_\tau = 1$, and $g^{-1}(x, q_\tau) = q_\tau - x^\intercal\beta$. Therefore:

$$\begin{aligned} UQPE_j(\tau; x) &= \beta_j \cdot \frac{f_\varepsilon(q_\tau - x^\intercal\beta)}{E[f_\varepsilon(q_\tau - X^\intercal\beta)]} \\ \text{and} \quad UQPE_j(\tau) &= \beta_j \end{aligned}$$

By definition of $UQPE_j$, β_j will then correspond to a unit increase in X_j on the marginal quantile of Y. In this very restricted model, the partial effects $UQPE_j(\tau; x)$ at different levels of X will be different and they will also be different along the marginal distribution of Y. However, averaging over the X the numerator and denominator cancels out and we will have that $UQPE_j = UAPE_j$. This is the case where not much is gained by considering unconditional or conditional quantiles, or alternatively it is a case that allows us to verify that the various parameters estimate what they promise to do.

4.3.2 Case 2: Non-linear, additively separable

Now, consider the special case of an index model $g(X, \varepsilon) = h(X^\intercal\beta + \varepsilon)$, where h is differentiable and monotonic. We keep the assumption that X and ε are independent. Then, $\partial g(X, q_\tau)/\partial X_j = \beta_j \cdot h'(X^\intercal\beta + q_\tau)$, $\partial g(x, q_\tau)/\partial q_\tau = h'(X^\intercal\beta + q_\tau)$, and $g^{-1}(x, q_\tau) = h^{-1}(q_\tau) - x^\intercal\beta$. The parameter of interest will be

$$\begin{aligned} UQPE_j(\tau; x) &= \beta_j \cdot f_\varepsilon\left(h^{-1}(q_\tau) - x^\intercal\beta\right) \cdot E\left[\frac{f_\varepsilon(h^{-1}(q_\tau) - X^\intercal\beta)}{h'(X^\intercal\beta + q_\tau)}\right]^{-1} \\ \text{and} \quad UQPE_j(\tau) &= \beta_j \cdot E\left[f_\varepsilon\left(h^{-1}(q_\tau) - X^\intercal\beta\right)\right] \cdot E\left[\frac{f_\varepsilon(h^{-1}(q_\tau) - X^\intercal\beta)}{h'(X^\intercal\beta + q_\tau)}\right]^{-1} \end{aligned}$$

Thus, the effect of small changes of X_j on the τ^{th} quantile of Y, $UQPE_j(\tau; x)$, will be heterogenous as it will in general depend on the value of x and the quantile itself. Here, after integrating X out according to its distribution, the partial effect, $UQPE_j(\tau)$ will depend on the quantile being evaluated. They will however be proportional to the structural parameters β_j.

With more structure, we are able to better establish the proportionality factor.

Adding a normality assumption about the distribution ε_i, that is, $\varepsilon_i \sim N(0,1)$, we can replace the f_ε by the normal density in the expression above. The marginal effects are then derived in the same way as in a Probit model

$$\begin{aligned} UQPE_j(\tau;x) &= \beta_j \cdot \phi\left(x^\intercal\beta - h^{-1}(q_\tau)\right) \cdot E\left[\frac{\phi\left(X^\intercal\beta - h^{-1}(q_\tau)\right)}{h'\left(X^\intercal\beta + q_\tau\right)}\right]^{-1} \\ UQPE_j(\tau) &= \beta_j \cdot E\left[\phi\left(X^\intercal\beta - h^{-1}(q_\tau)\right)\right] \cdot E\left[\frac{\phi\left(X^\intercal\beta - h^{-1}(q_\tau)\right)}{h'\left(X^\intercal\beta + q_\tau\right)}\right]^{-1} \end{aligned}$$

To see why this is similar to the marginal effects of a Probit, note that in this case, equation (10) simplifies to

$$UQPE_j(\tau;x) = \frac{dE\left[\mathrm{RIF}(Y;q_\tau)|X=x\right]}{dx_j} = \beta_j \cdot \frac{1}{f_Y(q_\tau)} \cdot \phi\left(x^\intercal\beta - h^{-1}(q_\tau)\right)$$

where

$$\begin{aligned} f_Y(q_\tau) &= \frac{d\Pr\left[Y \le q_\tau\right]}{dq_\tau} = \frac{dE\left[\Pr\left[Y \le q_\tau|X\right]\right]}{dq_\tau} \\ &= \frac{dE\left[\Phi\left(-X^\intercal\beta + h^{-1}(q_\tau)\right)\right]}{dq_\tau} = E\left[\frac{\phi\left(X^\intercal\beta - h^{-1}(q_\tau)\right)}{h'\left(X^\intercal\beta + q_\tau\right)}\right] \end{aligned}$$

Since the marginal effects in a probit model tend to be quite close to the slope estimates in a linear probability model, this gives a nice interpretation of the recentered influence function linear projection coefficients.[14]

4.3.3 Case 3: Non-linear, additively separable, Gaussian unobservables, Heteroscedastic Probit

Consider an alternative model where the error terms are possibly heteroscedastic, i.e. $Var(\varepsilon_i|X_i) = \sigma^2(x_i) \neq \sigma^2$. Because this is a violation of the independence assumption, we use the first result $i)$ of Proposition 1. We will also assume that $g(X,\varepsilon) = h\left(X^\intercal\beta + \varepsilon\right)$, where $\varepsilon_i = \sigma(x_i) \cdot \eta_i$, where $\eta|X \sim \eta \sim N(0,1)$. Thus, we have:

$$F_{\varepsilon|X}\left(g^{-1}(x,q_\tau)|x\right) = F_{\varepsilon|X}\left(h^{-1}(q_\tau) - x^\intercal\beta|x\right) = \Phi\left(\frac{x^\intercal\beta - h^{-1}(q_\tau)}{\sigma(x)}\right)$$

[14]We note that while software such as STATA computes the marginal effects of the probit model at the mean of the covariates, Wooldridge (2002) for example suggests taking the average valued of the individual marginal effects as an alternative.

and

$$f_{\varepsilon|X}\left(g^{-1}\left(x,q_\tau\right)\right)=\frac{1}{\sigma\left(x\right)}\cdot\phi\left(\frac{x^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma\left(x\right)}\right)$$

and

$$\frac{\partial F_{\varepsilon|X}\left(\cdot|x\right)}{\partial x_j}=-\frac{\partial\sigma\left(x\right)}{\partial x_j}\cdot\left(\frac{x^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma^2\left(x\right)}\right)\cdot\phi\left(\frac{x^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma\left(x\right)}\right)$$

Therefore:

$$\begin{aligned}UQPE_j\left(\tau;x\right) &= \phi\left(\frac{x^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma\left(x\right)}\right)\cdot\left(\frac{\beta_j}{\sigma\left(x\right)}+\frac{\partial\sigma\left(x\right)}{\partial x_j}\cdot\left(\frac{x^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma^2\left(x\right)}\right)\right)\\ &\quad\cdot E\left[\frac{1}{\sigma\left(X\right)}\cdot\phi\left(\frac{X^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma\left(X\right)}\right)\Big/h'\left(X^{\intercal}\beta+q_\tau\right)\right]^{-1}\end{aligned}$$

and

$$\begin{aligned}UQPE_j\left(\tau\right) &= E\left[\phi\left(\frac{X^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma\left(X\right)}\right)\cdot\left(\frac{\beta_j}{\sigma\left(X\right)}+\frac{\partial\sigma\left(X\right)}{\partial x_j}\cdot\left(\frac{X^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma^2\left(X\right)}\right)\right)\right]\\ &\quad\cdot E\left[\frac{1}{\sigma\left(X\right)}\cdot\phi\left(\frac{X^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma\left(X\right)}\right)\Big/h'\left(X^{\intercal}\beta+q_\tau\right)\right]^{-1}\end{aligned}$$

Because of the heteroscedasticity factor, the $UQPE_j(\tau)$ is no longer proportional to structural parameters, instead it is a weighed combination of the β_j and of the indirect effects of $\sigma(X)$. In the heteroscedastic case, the $UQPE_j(\tau)$ can even change sign with the quantile, if the second term in the expression

$$\left(\frac{\beta_j}{\sigma\left(X\right)}+\frac{\partial\sigma\left(X\right)}{\partial x_j}\cdot\left(\frac{X^{\intercal}\beta-h^{-1}\left(q_\tau\right)}{\sigma^2\left(X\right)}\right)\right)$$

dominates the first term and is of opposite sign. In the union example below, the impact of unionization will be positive for low τ and negative for high τ.

5 Estimation

In this section, our focus will be on the estimation of $UQPE\left(\tau\right)$, which is unsurpringly more involved than the case of the mean. We first have to estimate the recentered influence function, which will depend on some unknown objects (the quantile and the density) of the marginal distribution of Y. Then, in the estimation of parameters $\kappa\left(q_\tau;\xi\right)$, we will have to consider the estimation of $E\left[\text{RIF}(Y;q_\tau)|X=x\right]$ and its derivative, which will be important for the $UQPE\left(\tau\right)$ parameters.

The simplest case for the estimation of the expectation of the recentered influence function conditional of the X is a simple OLS regression $\widehat{E}\left[\widehat{\mathrm{RIF}}(Y;\widehat{q}_\tau)\ |X=x\right]=x'\cdot\widehat{\gamma}_\tau$, whose consistency will depend on the linearity of the recentered influence function at $X=x$. We call this method the RIF-OLS regression.

The results from section 3.2 indicate, however, that $dE\left[\mathrm{RIF}(Y;q_\tau)\ |X=x\right]/dx$ will generally depend on x, even when Y depends linearly on X and unobservables are independent of X. In order to deal with the fact that $dE\left[\mathrm{RIF}(Y;q_\tau)\ |X=x\right]/dx$ will depend on x, we consider an alternative nonparametric estimation procedure, which is nonparametric in the sense that it does not assume any functional form (as in RIF-OLS) or distribution for the errors (as in the probit marginal effects).

For a general functional of the marginal distribution of Y, $\nu\left(F_Y\right)$, we have that the parameters of interest would be $dE\left[\mathrm{RIF}(Y;\nu)|X=x\right]/dx$ and its expectation, $E\left[dE\left[\mathrm{RIF}(Y;\nu)|X\right]/dX\right]$. The last parameter is the expectation of a derivative and, for continuous X with unbounded support, the estimation of $E\left[dE\left[\mathrm{RIF}(Y;q_\tau)|X\right]/dX\right]$ will follow standard methods of average derivative estimation.[15]. In this paper we do not proceed with this possible way to estimate $UQPE\left(\tau\right)$ as in many applications we will need an estimator that has nice properties for covariates with bounded support and have high dimensionality.

Actually, for the particular case of quantiles we can build on the result in Equation (9)

$$UQPE\left(\tau\right)=-\frac{1}{f_Y\left(q_\tau\right)}\cdot\int F'_{Y|X}\left(q_\tau|X=x\right)\cdot dF_X\left(x\right)$$

and estimate nonparametrically the quantities $f_Y\left(q_\tau\right)$ and $F'_{Y|X}\left(q_\tau|X=x\right)$ by respectively $\widehat{f}_Y\left(\widehat{q}_\tau\right)$ and $\widehat{F'}_{Y|X}\left(\widehat{q}_\tau|X=x\right)$. In doing so, we will be able to estimate $UQPE\left(\tau\right)$ by

$$\widehat{UQPE}_{NP}\left(\tau\right)=-\frac{1}{N\cdot\widehat{f}_Y\left(\widehat{q}_\tau\right)}\cdot\sum_{i=1}^{N}\widehat{F'}_{Y|X}\left(\widehat{q}_\tau|X=X_i\right)$$

The remainder of this section is divided as following. In the first part we present an estimator for the recentered influence function of quantiles. We then present sufficient assumptions and prove that the estimator will be uniformly consistent for the true RIF.

[15]Under an appropriate set of sufficient assumptions, we can prove the asymptotic normality of the non-parametric estimator of $E\left[dE\left[\mathrm{RIF}(Y;\nu)|X\right]/dX\right]$. The proof of this result is a direct consequence of Theorem 3.1 in Härdle and Stoker (1989), after making an adjustment that takes into account the fact that we have to estimate first $\mathrm{RIF}(y;\nu)$ by $\widehat{\mathrm{RIF}}(Y_i;\widehat{\nu})$. Details of this result are available upon request from the authors. However, the non-parametric method is empirically applicable only in cases where the covariates have close to unbounded support and their dimensionality is reduced as in Deaton and Ng (1998).

The second part presents the OLS-RIF method to estimate $UQPE(\tau)$ and its large-sample properties. We then present details of the nonparametric method and its large sample properties.

5.1 Estimating the Recentered Influence Functions

In order to estimate either $UQPE(\tau; x)$ or $UQPE(\tau)$ we first have to obtain the estimated recentered influence functions. We propose the following

$$\widehat{\mathrm{RIF}}(Y; \widehat{q}_\tau) = \widehat{q}_\tau + \frac{\tau - \mathbb{1}\{Y \leq \widehat{q}_\tau\}}{\widehat{f}_Y(\widehat{q}_\tau)}$$

which also involves two unknown quantities to be estimated $\widehat{q}_\tau$ and $\widehat{f}_Y(\cdot)$.[16] The estimator of the population quantile τ of the marginal distribution of Y is $\widehat{q}_\tau$, the ordinary τ-th sample quantile, which can be represented, using Koenker and Basset (1978) as

$$\widehat{q}_\tau = \arg\min_q \sum_{i=1}^{N} (\tau - \mathbb{1}\{Y_i - q \leq 0\}) \cdot (Y_i - q)$$

The estimator of the density of Y is $\widehat{f}_Y(\cdot)$, the kernel density estimator. In the empirical section we propose using the Gaussian kernel with associated optimal bandwidth. The actual requirements for the kernel and for the bandwidth are described in the asymptotics section. Let $K_Y(z)$ be a kernel function and b_Y a positive scalar bandwidth, such that for a point y in the support of Y:

$$\widehat{f}_Y(y) = \frac{1}{N \cdot b_Y} \cdot \sum_{i=1}^{N} K_Y\left(\frac{Y_i - y}{b_Y}\right) \tag{12}$$

5.1.1 Uniform consistency of the estimated recentered influence functions

For all y in the support of Y, we establish conditions that guarantee that $\widehat{\mathrm{RIF}}(y; \widehat{q}_\tau) \xrightarrow{P} \mathrm{RIF}(y; q_\tau)$:

Assumption 1 [***Conditions on the distribution of Y***] *(i) $F_Y(\cdot)$ is absolutely continuous and differentiable over $y \in \mathbb{R}$ and $f_Y(y) = dF_Y(y)/dy$; (ii) $\int y \cdot f_Y(y) \cdot dy < \infty$; (iii) $f_Y(\cdot)$ is uniformly continuous; (iv) $\int |f_Y(y)| \cdot dy < \infty$; (v) $f_Y(\cdot)$ is three times differentiable with bounded third derivative in a neighborhood of y; (vi) for $\tau \in (0,1)$, the*

[16] This is unlike the case of the mean where $\widehat{\mathrm{RIF}}(Y; \overline{Y}) = Y$.

sets $\Upsilon_\tau = \inf_q \{F_Y(q) \geq \tau\}$, *are singletons and their elements* $q_\tau \in \Upsilon_\tau$ *satisfy* $q_\tau < \infty$ *and* $f_Y(q_\tau) > 0$.

Assumption 2 [***Kernel Function and Bandwidth***] *(i)* $K_Y(\cdot)$ *is a bounded real-valued function satisfying (a)* $\int K_Y(y) \cdot dy = 1$, *(b)* $\int |K_Y(y)| \cdot dy < \infty$, *(c)* $\int K_Y^2(y) \cdot dy < \infty$, *(d)* $\lim_{y \to \pm\infty} |y| \cdot |K_Y(y)| = 0$, *(e)* $\sup_y |K_Y(y)| < \infty$, *(f)* $K_Y(y) = \frac{1}{2\pi} \int \exp(-i \cdot t \cdot y) \cdot \phi(t) \cdot dt$, *where* $\phi(t)$ *is the absolutely integrable characteristic function of* $K_Y(\cdot)$*; (ii)* h_Y *is a bandwidth sequence satisfying* $h_Y = O\left(N^{-1/6}\right)$ *and, therefore,* $\lim_{N \to +\infty} h_Y = 0$ *and* $\lim_{N \to +\infty} N \cdot h_Y^2 = +\infty$.

Lemma 1 *Under assumptions 1 and 2* $\sup_y \left| \widehat{\text{RIF}}(y; \widehat{q}_\tau) - \text{RIF}(y; q_\tau) \right| \xrightarrow{P} 0$

5.2 OLS-RIF Regression

In this subsection we propose a simple way to estimate the parameters we derived previously by means of OLS regressions. Like in the familiar OLS regression, we implicilty assume that the recentered influence function is linear in the covariates X, which may however include higher order or non-linear transformations of the original variables. If the linearity assumption seems inappropriate in particular applications, one can always turn to non-parametric estimation as proposed next. Moreover, OLS is known to produce the linear function of covariates that minimizes the specification error.

Consider first estimation of $m_\tau(x) = E\left[\text{RIF}(Y; q_\tau) \middle| X = x\right]$:

$$\widehat{m}_{\tau,linear}(x) = x^\intercal \cdot \widehat{\gamma}_\tau$$

where

$$\widehat{\gamma}_\tau = \left(\sum_{i=1}^{N} X_i \cdot X_i^\intercal \right)^{-1} \cdot \sum_{i=1}^{N} X_i \cdot \widehat{\text{RIF}}(Y_i; \widehat{q}_\tau) \tag{13}$$

Note that since $\mathbb{1}\{Y \leq \widehat{q}_\tau\}$ is a dummy variable for whether a given observation i is below (or above) the τ^{th} sample quantile, the recentered influence function projections are closely related to linear probability model. As $\mathbb{1}\{Y \leq \widehat{q}_\tau\}$ is divided by $\widehat{f}_Y(\widehat{q}_\tau)$, the projection coefficients $\widehat{\gamma}_\tau$ (except for the constant) are simply equal to the coefficients in a linear probability model divided by the rescaling factor $\widehat{f}_Y(\widehat{q}_\tau)$. Note that in this parametric case,

$$\widehat{UQPE}_{OLS}(\tau) = \widehat{\gamma}_\tau = \widehat{UQPE}_{OLS}(\tau; x).$$

5.2.1 Large-sample properties of $\widehat{UQPE}_{OLS}$

We start defining

$$U_{i,\tau} = \text{RIF}(Y_i; q_\tau) - X_i^\intercal \cdot \gamma_\tau$$

where

$$\gamma_\tau = \left(E\left[X \cdot X^\intercal\right]\right)^{-1} \cdot E\left[X \cdot \text{RIF}(Y; q_\tau)\right]$$

therefore, by definition of $U_{i,\tau}$ and γ_τ

$$E\left[U_{i,\tau} \cdot X_i\right] = 0$$

We make the following assumptions regarding the joint distribution of $U_{i,\tau}$ and X_i.

Assumption 3 [***Conditions on the joint distribution of X and U_τ***] *(i) $E\left[X \cdot X^\intercal\right]$ is invertible; (ii) all moments up to the fourth moment of X exist and are finite; and (iii) the variance of the product $U_\tau \cdot X$, $V\left[U_\tau \cdot X\right]$.*

Theorem 2 [***Asymptotic Normality of OLS estimator***] *Under the Assumptions invoked in Lemma 1 plus Assumption 3, letting X contain the constant, and with a random sample of (Y,X):*

$$\sqrt{N} \cdot \left(\widehat{\gamma}_\tau - \gamma_\tau\right) = \left(E\left[X \cdot X^\intercal\right]\right)^{-1} \cdot \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \cdot U_{i,\tau} + o_p\left(1\right)$$

where

$$\left(E\left[X \cdot X^\intercal\right]\right)^{-1} \cdot \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \cdot U_{i,\tau} \xrightarrow{D} N\left(0, \left(E\left[X \cdot X^\intercal\right]\right)^{-1} \cdot E\left[U_\tau^2 \cdot X \cdot X^\intercal\right] \cdot \left(E\left[X \cdot X^\intercal\right]\right)^{-1}\right)$$

5.3 Nonparametric Estimation of $UQPE\left(\tau\right)$

We will estimate $UQPE\left(\tau\right)$ nonparametrically by

$$\widehat{UQPE}_{NP}\left(\tau\right) = -\frac{1}{N \cdot \widehat{f}_Y\left(\widehat{q}_\tau\right)} \cdot \sum_{i=1}^{N} \widehat{F'}_{Y|X}\left(\widehat{q}_\tau | X = X_i\right)$$

The estimator for the density $\widehat{f}_Y\left(\cdot\right)$ and the sample quantile $\widehat{q}_\tau$ were already presented. We now show how to estimate $\widehat{F'}_{Y|X}\left(\widehat{q}_\tau | X = X_i\right)$. Then we discuss the asymptotic properties of our estimator.

Define

$$T_\tau \equiv \mathbb{1}\{Y \leq q_\tau\}$$

and the probability of $T_\tau = 1$ given $X = x$ is

$$F_{Y|X}\left(q_\tau|X = x\right) = \Pr\left[T_\tau = 1|X = x\right] \equiv p_\tau\left(x\right)$$

Suppose that q_τ was known. A nonparametric estimator for $\Pr\left[T_\tau = 1|X = x\right]$ was introduced by Hirano, Imbens and Ridder (2003). They propose using a polynomial series approximation to $\log\left(p_\tau\left(x\right)\right) - \log\left(1 - p_\tau\left(x\right)\right)$, i.e., to use polynomial functions of X to approximate the log odds ratio of $p_\tau\left(x\right)$.[17] The coefficients corresponding to those functions are estimated by a pseudo-maximum likelihood method.

Start by defining $H_{K(\tau)}(x) = [H_{K(\tau),j}(x)] \quad (j = 1, ..., K\left(\tau\right))$, a vector of length $K\left(\tau\right)$ of polynomial functions of $x \in \mathcal{X}$ satisfying the following properties: (i) $H_{K(\tau)} : \mathcal{X} \rightarrow \mathbb{R}^{K(\tau)}$; (ii) $H_{K(\tau),1}(x) = 1$, and (iii) if $K\left(\tau\right) > \left(n+1\right)^r$, then $H_{K(\tau)}(x)$ includes all polynomials up order n.[18] In what follows, we will assume that $K\left(\tau\right)$ is a function of the sample size N such that $K\left(\tau\right) \rightarrow \infty$ as $N \rightarrow \infty$.[19]

For a given value x of X, we define an unfeasible estimator of $p_\tau\left(x\right)$ by $\widehat{p}^U_{K(\tau)}(x) \equiv L(H_{K(\tau)}(x)^\intercal \widetilde{\pi}_{K(\tau)})$, where $L : \mathbb{R} \rightarrow \mathbb{R}$, $L(z) = (1 + exp(-z))^{-1}$; and

$$\widetilde{\pi}_{K(\tau)} = \arg\max_{\pi \in \mathbb{R}^{K(\tau)}} Q_N\left(\pi\right)$$

where

$$Q_N\left(\pi\right) = \sum_{i=1}^{N}\left(T_{\tau,i} \cdot \log(L(H_{K(\tau)}(X_i)^\intercal \pi)) + (1 - T_{\tau,i}) \cdot \log(1 - L(H_{K(\tau)}(X_i)^\intercal \pi))\right)$$

Such estimator is unfeasible as we do not know the true value q_τ. If instead we use the sample quantile $\widehat{q}_\tau$, we can define

$$\widehat{T}_\tau \equiv \mathbb{1}\{Y \leq \widehat{q}_\tau\}$$

[17] The log odds ratio of $p_\tau\left(x\right)$ is equal to $\log\left(p_\tau\left(x\right) / \left(1 - p_\tau\left(x\right)\right)\right)$.

[18] Further details regarding the choice of $H_{K(\tau)}(x)$ and its asymptotic properties can be found in Hirano, Imbens and Ridder (2003).

[19] Some criterion should be used in order to choose the length $K\left(\tau\right)$ as function of the sample size. For example, one could use a cross-validation method to choose the order of the polynomial.

and the feasible estimator of $p_\tau(x)$ will be $\widehat{p}^F_{K(\tau)}(x) \equiv L(H_{K(\tau)}(x)'\widehat{\pi}_{K(\tau)})$, where

$$\widehat{\pi}_{K(\tau)} = \arg \max_{\pi \in \mathbb{R}^{K(\tau)}} \widehat{Q}_N(\pi)$$

where

$$\widehat{Q}_N(\pi) = \sum_{i=1}^{N} \left(\widehat{T}_{\tau,i} \cdot \log(L(H_{K(\tau)}(X_i)^\intercal \pi)) + (1 - \widehat{T}_{\tau,i}) \cdot \log(1 - L(H_{K(\tau)}(X_i)^\intercal \pi)) \right)$$

Finally, the feasible estimator $\widehat{F'}_{Y|X}(\widehat{q}_\tau|X = x)$ is the derivative of $\widehat{p}^F_{K(\tau)}(x)$ with respect to x. This is:

$$\begin{aligned} \widehat{F'}_{Y|X}(\widehat{q}_\tau|X = x) &= \frac{dL\left(H_{K(\tau)}(x)^\intercal \widehat{\pi}_{K(\tau)}\right)}{dx} \\ &= \left(\frac{dH_{K(\tau)}(x)}{dx}\right)^\intercal \cdot \widehat{\pi}_{K(\tau)} \cdot L'\left(H_{K(\tau)}(x)^\intercal \widehat{\pi}_{K(\tau)}\right) \\ &= G_{K(\tau)}(x)^\intercal \cdot \widehat{\pi}_{K(\tau)} \cdot \widehat{p}^F_{K(\tau)}(x) \cdot \left(1 - \widehat{p}^F_{K(\tau)}(x)\right) \end{aligned}$$

where $G_{K(\tau)} : \mathbb{R}^k \to \mathbb{R}^{K(\tau) \times k}$, $G_{K(\tau)}(x) = dH_{K(\tau)}(x)/dx$ is the matrix $K(\tau) \times k$ of derivatives of the vector $H_{K(\tau)}(x)$ of polynomial functions of x with respect to x. Finally, we can write $\widehat{UQPE}_{NP}$ as

$$\widehat{UQPE}^\intercal_{NP} = -\frac{\widehat{\pi}^\intercal_{K(\tau)}}{N \cdot \widehat{f}_Y(\widehat{q}_\tau)} \cdot \sum_{i=1}^{N} G_{K(\tau)}(X_i) \cdot \widehat{p}^F_{K(\tau)}(X_i) \cdot \left(1 - \widehat{p}^F_{K(\tau)}(X_i)\right) \tag{14}$$

5.3.1 Large-sample properties of $\widehat{UQPE}_{NP}$

In order to derive the asymptotic distribution of $\widehat{UQPE}_{NP}$, we use the fact that $\widehat{UQPE}_{NP} = -\frac{1}{\widehat{f}_Y(\widehat{q}_\tau)} \cdot \sum_{i=1}^N \widehat{F'}_{Y|X}(\widehat{q}_\tau|X = X_i)/N$, that is, we have an average derivative estimator divided by $-\widehat{f}_Y(\widehat{q}_\tau)$. Under the assumptions invoked for Lemma 1, this scale factor will converge uniformly in probability to $-f_Y(q_\tau)$. Hence, all we need to compute is the limiting distribution of $\sum_{i=1}^N \widehat{F'}_{Y|X}(\widehat{q}_\tau|X = X_i)/N$.

Consider first the unfeasible estimator

$$\frac{1}{N} \cdot \sum_{i=1}^{N} \widehat{F'}_{Y|X}(q_\tau|X = X_i) = \frac{1}{N} \cdot \sum_{i=1}^{N} G_{K(\tau)}(X_i)^\intercal \cdot \widetilde{\pi}_{K(\tau)} \cdot \widehat{p}^U_{K(\tau)}(X_i) \cdot \left(1 - \widehat{p}^U_{K(\tau)}(X_i)\right)$$

of $E\left[F'_{Y|X}(q_\tau|X)\right] = E\left[\dot{p}_\tau(X)\right]$ where $\dot{p}_\tau(X) = \frac{dp_\tau(x)}{dx}|_{x=X}$ .Such estimator will be unfeasible since q_τ is unknown.

We now follow Newey (1994) and Newey and Stoker (1993) that established that regardless of choice of the estimation procedure in the first step, the average derivative estimators will have the same influence function.

Proposition 2 [***Asymptotic Distribution of Average Derivative Estimator***]: *Under the regularity conditions in Newey and Stoker (1993, Assumptions 3.1 and 3.2) we have*

$$\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{F'}_{Y|X} \left(q_\tau | X = X_i\right) - E\left[F'_{Y|X}\left(q_\tau|X\right)\right]\right)$$

$$= \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(F'_{Y|X} \left(q_\tau | X = X_i\right) - E\left[F'_{Y|X}\left(q_\tau|X\right)\right]\right)$$

$$+ \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} l\left(X_i\right) \cdot \left(T_{\tau,i} - p_\tau\left(X_i\right)\right) + o_p\left(1\right)$$

where $l(x) = d \ln f_X(x)/dx$.

We do not provide a proof of Proposition 2 as it is a special case of Newey and Stoker's (1993) Theorem 3.1. Note that the key difference from them is about the choice of the nonparametric estimator of $p_\tau(x)$ and its derivative, which does not affect the limiting distribution of the estimator. In their paper, Newey and Stoker consider estimating the conditional expectation by series. We exploit the fact that our conditional expectation is a conditional probability which is in $[0, 1]$ and estimate its log-odds ratio by series. Such difference should not affect the limiting distribution of the final estimator.

In reality we use the feasible estimator, which uses $\widehat{q}_\tau$ instead of q_τ. We now show that under the assumptions invoked in Lemma 1 and assumptions about the series approximation, this is also irrelevant for the limiting distribution, that is, the feasible and the unfeasible estimators will be asymptotically equivalent.

Proposition 3 [***Asymptotic Equivalence of Feasible and Unfeasible Estimators***]: *Under the assumptions for the Sieve-series approximation in Hirano, Imbens and Ridder*

(2003, Assumption 5), and assumptions 1 and 2 we have

$$\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{F'}_{Y|X} \left(\widehat{q}_\tau | X = X_i \right) - E \left[F'_{Y|X} \left(q_\tau | X \right) \right] \right)$$

$$= \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(F'_{Y|X} \left(q_\tau | X = X_i \right) - E \left[F'_{Y|X} \left(q_\tau | X \right) \right] \right)$$

$$+ \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} l \left(X_i \right) \cdot \left(T_{\tau,i} - p_\tau \left(X_i \right) \right) + o_p \left(1 \right)$$

Finally, a combination of the previous results allows us to establish that

Corollary 1 [***Asymptotic Distribution of*** $\widehat{UQPE}_{NP}$]*: Under the assumptions in propositions 2 and 3 we have that*

$$\sqrt{N} \cdot \left(\widehat{UQPE}_{NP} - UQPE \right) = -\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\frac{F'_{Y|X} \left(q_\tau | X = X_i \right)}{f_Y \left(q_\tau \right)} + UQPE \right) - \frac{1}{f_Y \left(q_\tau \right) \cdot \sqrt{N}} \cdot \sum_{i=1}^{N} l \left(X_i \right) \cdot \left(T_{\tau,i} - p_\tau \left(X_i \right) \right) + o_p \left(1 \right)$$

6 Empirical Applications

To illustrate how our methodology works in practice, we present two empirical applications. The first one revisits the birthweight model of Koencker and Hallock (2001), where the authors show that there are differential impacts of being a boy or having a black mother, for example, at different quantiles of the conditional birthweight distribution. While on average boys are larger than girls, the disparity is much smaller in the lower quantiles and considerably larger in the upper quantiles. The second empirical application focuses on the impact of unions on male wages which are well known to have differential impacts at different points in the wage distribution (e.g. Card (1996)). There are several reasons why the impact of unions may be different at different quantiles of the wage distribution. First, unions both increase the conditional mean of wages (the "between" effect) and decrease the conditional distribution of wages (the "within" effect). This means that unions tend to increase wages in low wage quantiles where both the between and within group effects go in the same direction, but can decrease wages in high wage quantiles where the between and within group effects go in opposite directions.

This is compounded by the fact that the union wage gap generally declines as a function of the (observed) skill level.

These two examples are also useful in illustrating differences between unconditional and conditional quantiles regressions. In the first case, the conditional and unconditional distribution of birthweight are very similar, both look like Gaussian distributions slightly shifted one from another, so that the unconditional and conditional quantiles regressions will yield estimates that are generally similar to each other.[20] Note also that despite a large sample of 198,377 observations, the standard errors are quite large, a pattern that can also be found in Figure 4 of Koencker and Hallock (2001).[21] The large standard errors mean that the covariates do very little to explain individual differences in birthweight, so it is not surprising that the conditional and unconditional quantile estimates are similar. These conditional and unconditional (using the RIF-OLS and the non-parametric estimation strategy) quantile estimates, are presented in Figure 1a. The point estimates are generally very close, even for the variables for which they appear different, given the large standard errors, displayed in Figure 1b, they are not statistically different.

In the second empirical application, however, the conditional and unconditional distribution of log wages are more dissimilar. For example, the distribution of log wages conditional on being covered by a union is not only shifted to the right of the unconditional distribution, but it is also a more compressed and skewed distribution, corresponding to location-scale-twist of the unconditional distribution. By contrast the distribution of wages for non-union workers is closer to a normal distribution, but it typically has a mass point in the lower tail at the minimum wage.[22] The conditional and unconditional quantile regression estimates for union status will thus be very different, as illustrated in Figures 2a and 2b. On the other hand, the distributions of log wages by different experience levels are less dissimilar and the corresponding unconditional and conditional quantile regression estimates will be closer. We now discuss in more detail the estimation results in the case of unions.

6.1 Unions and Wage Inequality

Table 1 reports the estimates using the RIF-OLS method at the 10^{th}, 50^{th} and 90^{th} quantiles using a large sample of U.S. males from the 1983-85 Outgoing Rotation group

[20]See the figures in Appendix Figure A1.

[21]We use the same June 1997 Detailed Natality Data published by the National Center for Health Statistics as used by Koencker and Hallock (2001).

[22]At least in 1983-85, see Appendix Figure A2.

(ORG) supplement of the Current Population Survey. We start with 1983 because it is the first year in which the ORG supplement asked about union status. The dependent variable is the real log hourly wage for all wage and salary workers. Other data processing details can be found in Lemieux (2006b). The results are also compared with the OLS benchmark, and with standard quantile regressions at the corresponding quantiles. Interestingly, the effect of unions first increases from 0.198 at the 10th quantile to 0.349 at the median before turning negative (-0.137) at the 90^{th} quantile. These findings strongly confirm the well known results that unions have different effects ($UQPE$ here, to be precise) at different points of the wage distribution. Note that the effects are very precisely estimates for all specifications, thanks to the large available sample sizes (266,956 observations).

The quantile regression estimates reported in the corresponding columns show, as in Chamberlain (1994), that unions increase the location of the conditional wage distribution (i.e. positive effect on the median) but also reduce conditional wage dispersion. This explains why the effect of unions monotonically declines from 0.288, to 0.195 and 0.088 as quantiles increase. One cannot infer from the quantile regressions, however, what is the overall effect of unions on the unconditional wage distribution. The key problem is that, unlike conditional means, conditional quantiles do not aggregate up to unconditional quantiles. For example, the fact that unions increase the conditional median by 0.196 does not say anything about the effect of unions on the unconditional median. In fact, Table 1 shows that the effect of unions on unconditional quantiles estimated using RIF-OLS regressions are quite different from the conditional quantile estimates. For instance, the effect on the median (0.349) largely exceeds the conditional quantile regression estimate of 0.195.

The difference between RIF-OLS estimates of $UQPE$ and quantile regression estimates of $CQPE$ is illustrated in detail in Figure 2a, which plots both sets of estimates for each covariate at 19 different quantiles (from the 5^{th} to the 95^{th}). The pattern of estimated union effect on conditional and unconditional quantiles is very different. The unconditional effect is highly non-monotonic, while the conditional effect declines monotonically. In particular, the unconditional effect first increases from about 0.1 at the 5^{th} quantile to about 0.4 at the 35^{th} quantile, before declining and eventually reaching a large negative effect of over -0.2 at the 95^{th} quantile. The large effect at the top end reflects the fact that compression effects dominate everything else there. Union wages are more compressed and do not exhibit the "fat" upper tail of the non-union wage distribution. As a result, unions have large and negative impact on the probability of earning more than the 95^{th} quantile.

By contrast, traditional (conditional) quantile regression estimates decline almost linearly from about 0.3 at the 5^{th} quantile, to barely more than 0 at the 95^{th} quantile. So while both conditional and unconditional quantile regressions generally indicate that unions reduce wage inequality, the implied impacts are very different at different points of the wage distribution. For instance, one may read the quantile regressions estimates as indicating that de-unionization would uniformly reduce inequality at different points of the distribution. Unconditional regressions show, however, that the effect is much more non-linear. For instance, since the impact on the 10th and 50th quantiles are quite similar, this suggests that a decline in unionization would not result in expanding "low-end" wage inequality measured by the 50-10 gap. But since unions have a much larger impact at the median than at the 90th percentile, this suggests that a decline in unionization would result in a large expansion in the 90-50 gap. This is an important finding since recent studies such as Lemieux (2006a) and Autor et al. (2006) show that inequality has expanded much more at the "top-end" than at the "low-end" over the last twenty years.

The difference between conditional and unconditional effects is also quite important for several other covariates. For example, the effect of being a high-school dropout (relative to a high school graduate) goes in opposite directions for conditional and unconditional regressions above the 15^{th} quantile. On the one hand, the fact that the (negative) effect of being a dropout increase with quantiles for conditional quantile regressions simply reflects the heteroskedasticity in the data (residual dispersion increases with education). On the other hand, the opposite happens for unconditional regressions. The reason is that while being a dropout instead of a high school graduate has a big negative impact at low quantiles, the effect is much smaller at higher quantiles where having or not completed high school changes very little to the probability of earning a very high wage (like the 95^{th} percentile). One basically "needs" some college education to have any chance of earning these higher wages.

Another interesting feature of the unconditional quantile regression estimates is that, for all covariates, the estimated effects at the 10^{th} and especially the 5^{th} quantile tend to go abruptly towards zero. For example, the negative wage effect of being "non-white" hovers between 0.15 to 0.20 from the 15^{th} to the 50^{th} quantile, but drops to 0.12 at the 10^{th}, and only 0.05 at the 5^{th} quantile. The situation looks even more extreme for the effect of having very low experience (less than five years). The explanation for this puzzling pattern is that the 5^{th} is very close to the value of the minimum wage. Since the bottom of the distribution gets "bunched-up" at the minimum wage, factors like not being white and having little experience no longer have a negative effect in this part of the distribution. The unconditional wage regressions thus capture this important feature of the wage distribution that is mostly masked in the conditional quantile regressions.

Figure 2b report similar estimates for the more recent period (2003-05). While most estimates look qualitatively similar to those for 1983-85, a number of interesting new patterns also emerge. For example, there is now a more marked difference for the conditional and unconditional effects linked to education categories. In particular, the unconditional effect of having a post-graduate degree now increases much more steeply from the 5^{th} to the 95^{th} quantile than the conditional effect.

The remaining figures illustrate that, in a number of respects, the RIF-OLS regressions appear to be providing very robust estimates of the underlying parameter of interest, the $UQPE$. Figure 2c compares the RIF-OLS estimates to those obtained by computing the marginal effects from a logit regression with the same parametrization as the one used for the RIF-OLS (see Table 1). In most cases, the two sets of estimates are very close to each other. This confirms the "common wisdom" in empirical work that marginal effects from the linear probability model (RIF-OLS) are very similar to those from a logit or probit. In the case of unions, the last panel of Figure 2d shows that the confidence intervals obtained using the two methods are very close to each other. This is in sharp contrast with the very big difference in confidence intervals obtained when comparing RIF-OLS estimates with conditional quantile regressions in the first panel of Figure 2d. Note also that we use bootstrap standard errors for the logit marginal effects to also take account of the fact that the density (denominator in the RIF) is estimated. Accounting for this source of variability has very little impact on the confidence intervals because densities are very precisely estimated in our large sample.

The second panel of Figure 2d shows, however, that even if the density is precisely estimated, the choice of the bandwidth does matter for some of the unconditional quantile regression at the 15^{th}, 20^{th}, and 25^{th} quantiles. The problem is that there is a lot of heaping at \$5 and \$10 in this part of the wage distribution, which makes the kernel density estimates erratic when small bandwidths (0.02 or even 0.04) are used. The figure suggests it is better to oversmooth a bit the data with a larger bandwidth (0.06) even when the sample size is very large. Oversmoothing makes the estimates better behaved between the 15^{th} and the 25^{th} quantile, but has very little impact at other quantiles. Finally, the third panel of Figure 2d shows that using the fully nonparametric estimator (flexible NP in the graph) yields estimates that are virtually identical to those obtained with the parametric logit or the RIF-OLS.[23]

Having established that the RIF-OLS method works very well in practice, we return to an important potential limitation of our approach that has to do with the fact that the influence function is only a derivative. As a result, the $UQPE$ estimated by RIF-OLS (or non-parametric methods) is only accurate for infinitesimal changes in the covariates. For larger changes in the covariates, there is an approximation error. How large is the approximation error is, however, an empirical question.

To assess the importance of the approximation error, we conduct a small experiment looking at the effect of unions (in the 1983-85 data) but ignoring all other covariates. To predict the effect of changes in unionization using our approach, we run RIF-OLS regressions using only union status as explanatory variable. We then predict the value of the quantile at different levels of unionization by computing the predicted value of the RIF for different values of the unionization rate. The straight lines in Figures 4a to 4g show the result of this exercise for various changes in the unionization rate relative to the baseline rate (26.2 percent).

Since we only have a dummy covariate, it is also straightforward in this case to compute an "exact" effect of unionization by simply changing the proportion of union workers, and recomputing the various quantiles in this "reweighted" sample.[24] The resulting estimates are the diamonds reported in Figure 3a to 3g. Generally speaking, the RIF-OLS estimates are remarkably close to the "exact" estimates, even for large changes in unionization (plus or minus 10 percentage points). So while this is only a very special case, the results suggest that our approach is a very good approximation that works both for small and larger changes in the distribution of covariates.

The very last panel of Figure 3 (Figure 3h) repeats the same exercise for the variance. The advantage of the variance is that, unlike quantiles, it is possible to find a closed form expression for the effect of unions on the variance. More specifically, the well known analysis of variance formula implies that the overall variance is given by:

$$Var(w) = U \cdot \sigma_u^2 + (1 - U) \cdot \sigma_n^2 + U \cdot (1 - U) \cdot D^2$$

where U is the unionization rate, σ_u^2 (σ_n^2) is the variance within the union (non-union) sector, and D is the union wage gap. The effect of a change ΔU in the unionization rate

[23] We fully interact union status with all the other variables shown in Table 1 to get a "non-parametric" effect for unions.

[24] This can be viewed as a special case of Dinardo and Lemieux (1997)'s reweighting estimator of the effect of unions, where they perform a conditional reweighting where other covariates are also controlled for.

is simply:

$$\Delta Var(w) = \Delta U \cdot (\sigma_u^2 - \sigma_n^2) + \left[\Delta U(1-2U) - (\Delta U)^2\right] \cdot D^2$$

For an infinitesimal change in ΔU , the derivative of $Var(w)$ with respect to U is

$$\frac{dVar(w)}{dU} = (\sigma_u^2 - \sigma_n^2) + (1-2U) \cdot D^2$$

Using the derivative to do a first-order approximation of $\Delta Var(w)$ thus yields:

$$\widetilde{\Delta Var(w)} = \Delta U \cdot \left[(\sigma_u^2 - \sigma_n^2) + (1-2U) \cdot D^2\right]$$

It is easy to show that running a RIF-OLS regression for the variance and using it to predict the effect of changes in unionization on the variance yields the approximation $\widetilde{\Delta Var(w)}$ while the exact effect is $\Delta Var(w)$ from above. The approximation error is, thus, the second order term $\Delta Var(w) - \widetilde{\Delta Var(w)} = (\Delta U)^2 \cdot D^2$. It corresponds to the difference between the straight lines and the diamonds in Figure 3g. The diamonds are on a quadratic curve because of the second order term $(\Delta U)^2 \cdot D^2$, but the linear curve approximates the quadratic very well even for large changes in the unionization rate. In other words, the RIF-OLS approach yields very similar results compared to the analysis of variance formula that has been widely used in the literature. The fact that the approximation errors for both the quantiles and the variance are very small gives us great confidence that our approach can be used to generalize the distributional analysis of unions (or other factors) to any quantile of the unconditional distribution.

7 Conclusion

In this paper, we propose a new intuitive and computationally simple method for estimating unconditional quantile regressions. The method is based on the first order VOM appromixation of the unconditional quantile, which relies on the well-known influence function. In its simplest form, it consists of a linear projection of the recentered influence function of a given quantile onto the covariates. The method is useful complement to conditional quantile regressions when the parameter of policy interest has to do with the unconditional distribution.

We begin our exposition by contrasting unconditional quantile regressions with the popular conditional quantile regressions. As is well-known, an important problem with conditional quantile regressions is that they do not aggregate up. That is, the conditional quantile evaluated at the mean of the covariates is not equal to the unconditional quantile. In many contexts, where there is a need to apply a decomposition method, such as the Oaxaca-Blinder approach, this is an undesirable property. On the contrary, an unconditional quantile evaluated at the mean of the covariates will be equal to the unconditional quantile. Thus our method can be used to construct policy conterfactuals or decomposition of the unconditional quantile in the same way OLS is typically used for the mean.

We then proceed with a review of the basic concepts underlying our first order VOM approximation, which we call recentered influence function, of any functional of a given distribution of the dependent variable of interest, observed in the presence of covariates. We also provide a general result on the impact of changes in the distribution of the covariates, which can be written in general terms as mixtures of the original distribution and of a small change in the direction of a counterfactual distribution of desired changes. Our central result describes the impact of a small change in the distribution of covariates in the direction of the counterfactual distribution of desired changes on the functional of interest, as the integral of the expectation of the recentered influence function with respect to the changes in distribution of the covariates. Because this result applies to any counterfactual distribution, it gives us a powerful tool to consider commonly estimated partial effects, as well effects related to discrete changes linked to any policy of interest. Also, although we develop the details of our estimation strategies for quantiles, because this central result implies to any functional, our simplest estimation strategy would also work for other functionals, such as the variance or the Gini coefficient.

Our next step is to apply this general result to the case where the functional of interest is a given quantile of the unconditional distribution and consider two important specific cases of the impact of changes in the covariates. The leading case is the one where the counterfactual distribution of desired changes corresponds to providing a small increase in the value of the vector of covariates to everyone in the population. Its corresponds to the Unconditional Quantile Partial Effect (UQPE), whose counterpart in the case of the mean is shown to be the Unconditional Average Partial Effect (UAPE) (Wooldridge, 2004). We show that the UQPE of a particular covariate can be expressed as the average derivative of the conditional expectation of the recentered influence function with respect to that covariate. This expression can also be written as the average derivative of the conditional distribution of the dependent variable with respect the covariate of interest, rescaled by the inverse of the density of the dependent variable at the given quantile. These two key expressions for the UQPE provide us with the basic formulas underlying our estimation strategies. We also explore the links between the UQPE and a structural model that links the dependent variable to observed and unobserved covariates through a given functional form. We provide the corresponding expression in the general case and illustrate these links more precisely in three special cases that illustrate some simplifications, as well as potential pitfalls, when the unobservables are or are not independent of the observed covariates.

In the estimation section, we provide the details of our two estimation strategies for the alternative expressions of the UQPE, as well as the corresponding asymptotic results. As already mentionned, the first one simply consists in running an OLS regression of the recentered influence function of the unconditional quantile of the dependent variable on the explanatory variables and is called RIF-OLS. The second one consists in estimating non-parametrically the average derivative involved. Both strategies are relatively straightforward since the estimation of their various components relies on known estimators for the quantile (Koenker and Basset, 1978), for the non-parametric kernel density function, and for the log odds ratio of the conditional probability for a given quantile (Hirano, Imbens and Ridder, 2003).

Finally in the empirical section, we revisit two classic applications of quantile regression that illustrates well the differences between conditional and unconditional quantile regressions. In the first one, which restimates the birthweight model of Koencker and Hallock (2001), the two types of quantile regressions are very close. In the second one, which considers in particular the impact of unions of wages, the results are more strinkingly different. While Chamberlain (1984) had found that unions have a much larger effect at lower than higher conditional quantiles, we actually find a negative effect of unions at the highest quantile of the wage distribution.

8 Appendix

Proof of Theorem 1: Let us define define $H_{Y,t}$ as the following probability distribution over the support of Y:

$$F_{Y,t\cdot G_Y} = (1-t)\cdot F_Y + t\cdot G_Y$$

We have

$$
\begin{aligned}
F_{Y,t\cdot G_Y}(y) &= t\cdot(G_Y - F_Y)(y) + F_Y(y) \\
&= t\cdot(G_Y - F_Y) + F_Y \\
&= t\cdot\int F_{Y|X}(y|X=x)\cdot d(G_X - F_X)(x) \\
&\qquad\qquad + \int F_{Y|X}(y|X=x)\cdot dF_X(x) \\
&= \int F_{Y|X}(y|X=x)\cdot d(t\cdot(G_X - F_X) + F_X)(x) \\
&= \int F_{Y|X}(y|X=x)\cdot dF_{X,t\cdot G_X}(x) \\
&= F_{Y,t\cdot G_X}(y)
\end{aligned}
$$

The effect on the functional ν of the marginal distribution of Y of an infinitesimal change in the distribution of X from F_X towards G_X is defined as:

$$
\begin{aligned}
\frac{\partial\nu(F_{Y,t\cdot G_Y})}{\partial t}|_{t=0} &= \int \mathrm{IF}(y;\nu,F_Y)\cdot d(G_Y - F_Y)(y) \\
&= \int \mathrm{RIF}(y;\nu)\cdot d(G_Y - F_Y)(y) \\
&= \frac{\partial\nu(F_{Y,t\cdot G_X})}{\partial t}|_{t=0}
\end{aligned}
$$

Now, note that

$$
\begin{aligned}
\frac{dF_{Y,t\cdot G_X}(y)}{dy} &= d\frac{\int\int^y dF_{Y|X}(z|X=x)\cdot dF_{X,t\cdot G_X}(x)}{dy} \\
&= \int f_{Y|X}(y|X=x)\cdot dF_{X,t\cdot G_X}(x)
\end{aligned}
$$

thus,

$$
\begin{aligned}
\frac{\partial \nu\left(F_{Y,t\cdot G_X}\right)}{\partial t}|_{t=0} &= \int \mathrm{RIF}(y;\nu)\cdot d\left(G_Y - F_Y\right)(y) \\
&= \frac{1}{t}\cdot\int \mathrm{RIF}(y;\nu)\cdot dF_{Y,t\cdot G_X}(y) \\
&= \frac{1}{t}\cdot\int_y \mathrm{RIF}(y;\nu)\cdot\int_x f_{Y|X}(y|X=x)\cdot dF_{X,t\cdot G_X}(x)\cdot dy \\
&= \int_x\left(\int_y \mathrm{RIF}(y;\nu)\cdot f_{Y|X}(y|X=x)\cdot dy\right)\cdot\frac{dF_{X,t\cdot G_X}(x)}{t} \\
&= \int E\left[\mathrm{RIF}(Y;\nu)|X=x\right]\cdot d\left(G_X - F_X\right)(x)
\end{aligned}
$$

$\blacksquare$

Derivation of the Influence Function of a Quantile:
Let the τ^{th} quantile be defined implicitly as

$$
\tau = \int_{-\infty}^{q_\tau} dF(y) = \int_{-\infty}^{\nu(F)} dF(y) = \int_{-\infty}^{\nu(F_{t,y})} dF_{t,y}(y)
$$

Then by taking the derivative of the last expression with respect to t we obtain:

$$
\begin{aligned}
0 &= \frac{\partial\int_{-\infty}^{\nu(F_{t,y})} dF_{t,y}(y)}{\partial t} \\
&= \frac{\partial\nu\left(F_{t,y}\right)}{\partial t}\cdot\frac{dF_{t,y}(y)}{dy}|_{y=\nu(F_{t,y})} + \int_{-\infty}^{\nu(F_{t,y})} d\left(\Delta_y - F\right)(y) \\
&= \frac{\partial\nu\left(F_{t,y}\right)}{\partial t}\cdot\frac{dF_{t,y}(y)}{dy}|_{y=\nu(F_{t,y})} + \mathbb{1}\left\{y\le\nu(F_{t,y})\right\} - \int \mathbb{1}\left\{y\le\nu(F_{t,y})\right\} dF(y)
\end{aligned}
$$

Thus:

$$
\frac{\partial\nu\left(F_{t,y}\right)}{\partial t} = \frac{\int \mathbb{1}\left\{y\le\nu(F_{t,y})\right\} dF(y) - \mathbb{1}\left\{y\le\nu(F_{t,y})\right\}}{\frac{dF_{t,y}(y)}{dy}|_{y=\nu(F_{t,y})}}
$$

$\blacksquare$

Proof of Proposition 1:

$$
\begin{aligned}
UQPE_j(x) &= \frac{\partial E\left[\mathrm{RIF}(Y;q_\tau)|X=x\right]}{\partial x_j} \\
&= \partial\frac{\{q_\tau+(\tau-E\left[\mathbb{1}\{g(x,\varepsilon)\le q_\tau\}|X=x\right])/f_Y(q_\tau)\}}{\partial x_j} \\
&= \partial\frac{\left\{q_\tau+(\tau-E\left[\mathbb{1}\{\varepsilon\le g^{-1}(x,q_\tau)\}|X=x\right])/f_Y(q_\tau)\right\}}{\partial x_j} \\
&= -\left(f_Y\left(q_\tau\right)\right)^{-1}\cdot\frac{\partial F_{\varepsilon|X}\left(g^{-1}\left(x,q_\tau\right)|x\right)}{\partial x_j} \\
&= -\left(f_Y\left(q_\tau\right)\right)^{-1}\cdot\left(\frac{\partial g^{-1}\left(x,q_\tau\right)}{\partial x_j}\cdot f_{\varepsilon|X}\left(g^{-1}\left(x,q_\tau\right)|x\right)+\frac{\partial F_{\varepsilon|X}\left(\cdot|x\right)}{\partial x_j}\right)
\end{aligned}
$$

where defining the function H as

$$
H\left(g^{-1}\left(\cdot\right),x_1,x_2,...,x_j,...,x_k\right)\equiv F_{\varepsilon|X}\left(g^{-1}\left(x,q_\tau\right)|x\right)
$$

$$
\Rightarrow \frac{\partial F_{\varepsilon|X}\left(\cdot|x\right)}{\partial x_j}=\frac{\partial H\left(g^{-1}\left(\cdot\right),x_1,x_2,...,x_j,...,x_k\right)}{\partial x_j}
$$

Let us now work out an expression for the density of Y at q_τ:

$$
\begin{aligned}
f_Y\left(q_\tau\right) &= \frac{d\Pr\left[Y\le q_\tau\right]}{dq_\tau}=\frac{d\Pr\left[\varepsilon\le g^{-1}\left(X,q_\tau\right)\right]}{dq_\tau} \\
&= \frac{d\int\Pr\left[\varepsilon\le g^{-1}\left(x,q_\tau\right)|X=x\right]\cdot dF_X\left(x\right)}{dq_\tau} \\
&= \int\frac{d\Pr\left[\varepsilon\le g^{-1}\left(x,q_\tau\right)|X=x\right]}{dq_\tau}\cdot dF_X\left(x\right) \\
&= \int\frac{\partial g^{-1}\left(x,q_\tau\right)}{\partial q_\tau}\cdot f_{\varepsilon|X}\left(g^{-1}\left(x,q_\tau\right)|x\right)\cdot dF_X\left(x\right) \\
&= \int\frac{f_{\varepsilon|X}\left(g^{-1}\left(x,q_\tau\right)|x\right)}{\partial g\left(x,q_\tau\right)/\partial q_\tau}\cdot dF_X\left(x\right)
\end{aligned}
$$

and an expression for $\partial g^{-1}\left(x,q_\tau\right)/\partial x_j$, which uses the implicit function theorem:

$$
\frac{\partial g^{-1}\left(x,q_\tau\right)}{\partial x_j}=-\frac{\partial g\left(x,q_\tau\right)/\partial x_j}{\partial g\left(x,q_\tau\right)/\partial q_\tau}
$$

therefore, we have that $UQPE_j(x)$ can be expressed as:

$$\begin{aligned} UQPE_j(x) &= (f_Y(q_\tau))^{-1} \cdot \left(\frac{\partial g^{-1}(x, q_\tau)}{\partial x_j} \cdot f_{\varepsilon|X}\left(g^{-1}(x, q_\tau)\,|x\right) + \frac{\partial F_{\varepsilon|X}(\cdot|x)}{\partial x_j} \right) \\ &= \left(\frac{\partial g(x, q_\tau)}{\partial x_j} \cdot \frac{f_{\varepsilon|X}(g^{-1}(x, q_\tau))}{\partial g(x, q_\tau)/\partial q_\tau} - \frac{\partial F_{\varepsilon|X}(\cdot|x)}{\partial x_j} \right) \\ &\quad \cdot \left(\int \frac{f_{\varepsilon|X}(g^{-1}(x, q_\tau))}{\partial g(x, q_\tau)/\partial q_\tau} \cdot dF_X(x) \right)^{-1} \end{aligned}$$

Now, let us use the fact that ε and X are independent, that is, $\partial F_{\varepsilon|X}(\cdot|x)/\partial x_j = 0$ and $f_{\varepsilon|X}(\cdot) = f_\varepsilon(\cdot)$:

$$UQPE_j(x) = \frac{\partial g(x, q_\tau)}{\partial x_j} \cdot \frac{f_\varepsilon(g^{-1}(x, q_\tau))}{\partial g(x, q_\tau)/\partial q_\tau} \Big/ \int \frac{f_\varepsilon(g^{-1}(x, q_\tau))}{\partial g(x, q_\tau)/\partial q_\tau} \cdot dF_X(x)$$

■

Proof of Lemma 1:
Start by opening up the expression $\sup_y \left|\widehat{\mathrm{RIF}}(y; \widehat{q}_\tau) - \mathrm{RIF}(y; q_\tau)\right|$:

$$\begin{aligned} &\sup_y \left|\widehat{\mathrm{RIF}}(y; \widehat{q}_\tau) - \mathrm{RIF}(y; q_\tau)\right| \\ =\ &\sup_y \left| \widehat{q}_\tau + \frac{\tau - \mathbb{1}\{y \le \widehat{q}_\tau\}}{\widehat{f}_Y(\widehat{q}_\tau)} - q_\tau - \frac{\tau - \mathbb{1}\{y \le q_\tau\}}{f_Y(q_\tau)} \right| \end{aligned}$$

and that is bounded by

$$\begin{aligned} \le\ & |\widehat{q}_\tau - q_\tau| \\ & + \tau \cdot \left| \frac{\widehat{f}_Y(\widehat{q}_\tau) - f_Y(q_\tau)}{f_Y(q_\tau) \cdot \widehat{f}_Y(\widehat{q}_\tau)} \right| \\ & + \sup_y \left| \frac{\mathbb{1}\{y \le \widehat{q}_\tau\} - \mathbb{1}\{y \le q_\tau\}}{\widehat{f}_Y(\widehat{q}_\tau)} \right| \\ & + \sup_y \left| \left(\frac{\widehat{f}_Y(\widehat{q}_\tau) - f_Y(q_\tau)}{f_Y(q_\tau) \cdot \widehat{f}_Y(\widehat{q}_\tau)} \right) \cdot \mathbb{1}\{y \le q_\tau\} \right| \end{aligned}$$

Following for example van der Vaart (1998, p. 55), under Assumption 1:

$$\sqrt{N} \cdot (\widehat{q}_\tau - q_\tau) = \frac{\tau - 1\!\!1\{Y \leq q_\tau\}}{f_Y(q_\tau)} + o_p(1) \xrightarrow{D} N\left(0, \frac{\tau \cdot (1-\tau)}{f_Y^2(q_\tau)}\right)$$

thus, Equation (15) is $O_p\left(N^{-1/2}\right)$, that is, there is a constant c, such that, with high probability

$$|\widehat{q}_\tau - q_\tau| < \frac{c}{\sqrt{N}} = O\left(N^{-1/2}\right)$$

Equation (15) can be bounded in probability by noticing that:

$$\begin{aligned}
&\tau \cdot \left| \frac{\widehat{f}_Y(\widehat{q}_\tau) - f_Y(q_\tau)}{f_Y(q_\tau) \cdot \widehat{f}_Y(\widehat{q}_\tau)} \right| \\
\leq\ & c_1 \cdot \frac{\left(\left| \widehat{f}_Y(\widehat{q}_\tau) - f_Y(\widehat{q}_\tau) \right| + |f_Y(\widehat{q}_\tau) - f_Y(q_\tau)| \right)}{f_Y(q_\tau) \cdot \left| \left(\widehat{f}_Y(\widehat{q}_\tau) - f_Y(\widehat{q}_\tau) + f_Y(\widehat{q}_\tau) - f_Y(q_\tau) + f_Y(q_\tau) \right) \right|} \\
\leq\ & c_1 \cdot \frac{\left(\sup_y \left| \widehat{f}_Y(y) - f_Y(y) \right| + |f_Y(\widehat{q}_\tau) - f_Y(q_\tau)| \right)}{f_Y(q_\tau) \cdot \left(f_Y(q_\tau) - c_2 \cdot \sup_y \left| \widehat{f}_Y(y) - f_Y(y) \right| - c_3 \cdot |f_Y(\widehat{q}_\tau) - f_Y(q_\tau)| \right)}
\end{aligned}$$

where c_1, c_2 and c_3 are constants. By Assumption 1, $f_Y(y)$ is and continuously differentiable for all y in the support of Y, which allows us to apply the so-called Delta-method:

$$\sqrt{N} \cdot (f_Y(\widehat{q}_\tau) - f_Y(q_\tau)) = f'_Y(q_\tau) \cdot \sqrt{N} \cdot (\widehat{q}_\tau - q_\tau) + o_p(1) \xrightarrow{D} N\left(0, (f'_Y(q_\tau))^2 \cdot \frac{\tau \cdot (1-\tau)}{f_Y^2(q_\tau)}\right)$$

It is important here to invoke the work out a known result (for example, Pagan and Ullah, 1999, Theorems 2.4 and 2.8 on pp. 33-37) that follows from assumptions 1 and 2 about the kernel density estimator $\widehat{f}_Y(\cdot)$ being uniformly consistent to $f_Y(\cdot)$:

$$\begin{aligned}
\sup_y \left| \widehat{f}_Y(y) - f_Y(y) \right| &\leq \sup_y \left| \widehat{f}_Y(y) - E\left[\widehat{f}_Y(y)\right] \right| + \sup_y \left| E\left[\widehat{f}_Y(y)\right] - f_Y(y) \right| \\
&= O_p\left(\left(N \cdot h_Y^2\right)^{-1/2} \right) + O\left(h_Y^2\right) \\
&= O_p\left(N^{-1/2} \cdot N^{1/6}\right) + O\left(N^{-1/3}\right) = O_p\left(N^{-1/3}\right)
\end{aligned}$$

Therefore

$$\begin{aligned}
\tau \cdot \left| \frac{\widehat{f}_Y(\widehat{q}_\tau) - f_Y(q_\tau)}{f_Y(q_\tau) \cdot \widehat{f}_Y(\widehat{q}_\tau)} \right| &= O_p(1) \cdot \left(\frac{O_p\left(N^{-1/3}\right) + O_p\left(N^{-1/2}\right)}{O_p(1) + O_p\left(N^{-1/3}\right) + O_p\left(N^{-1/2}\right)} \right) = \frac{O_p\left(N^{-1/3}\right)}{O_p(1)} \\
&= O_p\left(N^{-1/3}\right)
\end{aligned}$$

Next we consider Equation (15)

$$\begin{aligned}
& \sup_y \left| \frac{\mathbb{1}\left\{y \leq \widehat{q}_\tau\right\} - \mathbb{1}\left\{y \leq q_\tau\right\}}{\widehat{f}_Y\left(\widehat{q}_\tau\right)} \right| \\
\leq \quad & \frac{\sup_y \left|\mathbb{1}\left\{y \leq \widehat{q}_\tau\right\} - \mathbb{1}\left\{y \leq q_\tau\right\}\right|}{f_Y\left(q_\tau\right) - c_1 \cdot \sup_y \left|\widehat{f}_Y\left(y\right) - f_Y\left(y\right)\right| - c_2 \cdot \left|f_Y\left(\widehat{q}_\tau\right) - f_Y\left(q_\tau\right)\right|}
\end{aligned}$$

Define the function $a_y\left(q\right) = \mathbb{1}\left\{q \geq y\right\}$, which is Lipschitz for all $y \in \mathbb{R}$, that is, for small real ϵ, there exist a constant c_y such that

$$\sup_y \left|a_y\left(q + \epsilon\right) - a_y\left(q\right)\right| < c_y \cdot \left|\epsilon\right|$$

and, therefore, there will be a constant c_1 that with high probability:

$$\begin{aligned}
& \sqrt{N} \cdot \sup_y \left|\mathbb{1}\left\{y \leq \widehat{q}_\tau\right\} - \mathbb{1}\left\{y \leq q_\tau\right\}\right| \\
\leq \quad & \sqrt{N} \cdot \sup_y \left|\mathbb{1}\left\{y \leq q_\tau + c_1/\sqrt{N}\right\} - \mathbb{1}\left\{y \leq q_\tau\right\}\right| \\
\leq \quad & \sqrt{N} \cdot c_y \cdot \left|c_1/\sqrt{N}\right| \\
= \quad & O\left(1\right)
\end{aligned}$$

Thus:

$$\begin{aligned}
& \sup_y \left| \frac{\mathbb{1}\left\{y \leq \widehat{q}_\tau\right\} - \mathbb{1}\left\{y \leq q_\tau\right\}}{\widehat{f}_Y\left(\widehat{q}_\tau\right)} \right| \\
= \quad & \frac{O_p\left(N^{-1/2}\right)}{O_p\left(1\right) + O_p\left(N^{-1/3}\right) + O_p\left(N^{-1/2}\right)} = O_p\left(N^{-1/2}\right)
\end{aligned}$$

Finally, consider Equation (15),

$$\begin{aligned}
& \sup_y \left| \left(\frac{\widehat{f}_Y\left(\widehat{q}_\tau\right) - f_Y\left(q_\tau\right)}{f_Y\left(q_\tau\right) \cdot \widehat{f}_Y\left(\widehat{q}_\tau\right)} \right) \cdot \mathbb{1}\left\{y \leq q_\tau\right\} \right| \\
\leq \quad & \left| \left(\frac{\widehat{f}_Y\left(\widehat{q}_\tau\right) - f_Y\left(q_\tau\right)}{f_Y\left(q_\tau\right) \cdot \widehat{f}_Y\left(\widehat{q}_\tau\right)} \right) \right| = O_p\left(N^{-1/3}\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sup_y \left|\widehat{\mathrm{RIF}}\left(y;\widehat{q}_\tau\right)-\mathrm{RIF}\left(y;q_\tau\right)\right| &= O_p\left(N^{-1/2}\right)+O_p\left(N^{-1/3}\right)+O_p\left(N^{-1/2}\right)+O_p\left(N^{-1/3}\right) \\
&= O_p\left(N^{-1/3}\right)=o_p\left(1\right)
\end{aligned}$$

■

Proof of Theorem 2: Now, let us rewrite $\widehat{\gamma}_\tau$

$$\begin{aligned}
\widehat{\gamma}_\tau &= \left(\sum_{i=1}^N X_i\cdot X_i^\intercal\right)^{-1}\cdot\sum_{i=1}^N X_i\cdot\widehat{\mathrm{RIF}}(Y_i;\widehat{q}_\tau) \\
&= \left(\sum_{i=1}^N X_i\cdot X_i^\intercal\right)^{-1}\cdot\sum_{i=1}^N X_i\cdot\mathrm{RIF}(Y_i;q_\tau) \\
&\quad+\left(\sum_{i=1}^N X_i\cdot X_i^\intercal\right)^{-1}\cdot\sum_{i=1}^N X_i\cdot\left(\widehat{\mathrm{RIF}}(Y_i;\widehat{q}_\tau)-\mathrm{RIF}(Y_i;q_\tau)\right) \\
&= \gamma_\tau+\left(\sum_{i=1}^N X_i\cdot X_i^\intercal\right)^{-1}\cdot\sum_{i=1}^N X_i\cdot U_{i,\tau} \\
&\quad+\left(\sum_{i=1}^N X_i\cdot X_i^\intercal\right)^{-1}\cdot\sum_{i=1}^N X_i\cdot\left(\widehat{\mathrm{RIF}}(Y_i;\widehat{q}_\tau)-\mathrm{RIF}(Y_i;q_\tau)\right)
\end{aligned}$$

Thus:

$$\begin{aligned}
&\left\|\sqrt{N}\cdot\left(\widehat{\gamma}_\tau-\gamma_\tau\right)-\left(E\left[X\cdot X^\intercal\right]\right)^{-1}\cdot\frac{1}{\sqrt{N}}\cdot\sum_{i=1}^N X_i\cdot U_{i,\tau}\right\| \\
\leq&\left\|\left(\left(\frac{1}{N}\cdot\sum_{i=1}^N X_i\cdot X_i^\intercal\right)^{-1}-\left(E\left[X\cdot X^\intercal\right]\right)^{-1}\right)\cdot\frac{1}{\sqrt{N}}\cdot\sum_{i=1}^N X_i\cdot U_{i,\tau}\right\| \qquad (15)\\
&+\left\|\left(\left(\frac{1}{N}\cdot\sum_{i=1}^N X_i\cdot X_i^\intercal\right)^{-1}-\left(E\left[X\cdot X^\intercal\right]\right)^{-1}\right)\cdot\frac{1}{\sqrt{N}}\cdot\sum_{i=1}^N X_i\cdot\left(\widehat{\mathrm{RIF}}(Y_i;\widehat{q}_\tau)-\mathrm{RIF}(Y_i;q_\tau)\right)\right\| \qquad (16)\\
&+\left\|\left(E\left[X\cdot X^\intercal\right]\right)^{-1}\cdot\frac{1}{\sqrt{N}}\cdot\sum_{i=1}^N X_i\cdot\left(\widehat{\mathrm{RIF}}(Y_i;\widehat{q}_\tau)-\mathrm{RIF}(Y_i;q_\tau)\right)\right\| \qquad (17)
\end{aligned}$$

Given Assumption 3, we can apply a central limit theorem to get

$$\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \cdot U_{i,\tau} \xrightarrow{D} N\left(0, E\left[U_\tau^2 \cdot X \cdot X^\intercal\right]\right)$$

and therefore

$$\left(E\left[X \cdot X^\intercal\right]\right)^{-1} \cdot \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \cdot U_{i,\tau} \xrightarrow{D} N\left(0, \left(E\left[X \cdot X^\intercal\right]\right)^{-1} \cdot E\left[U_\tau^2 \cdot X \cdot X^\intercal\right] \cdot \left(E\left[X \cdot X^\intercal\right]\right)^{-1}\right)$$

thus, $N^{-1/2} \cdot \sum_{i=1}^{N} X_i \cdot U_{i,\tau} = O_p(1)$. Also by Assumption 3 we can apply a law of large numbers to get

$$\left(\frac{1}{N} \cdot \sum_{i=1}^{N} X_i \cdot X_i^\intercal\right)^{-1} - \left(E\left[X \cdot X^\intercal\right]\right)^{-1} = o_p(1)$$

thus Equation (15) is

$$\left\| \left(\left(\frac{1}{N} \cdot \sum_{i=1}^{N} X_i \cdot X_i^\intercal\right)^{-1} - \left(E\left[X \cdot X^\intercal\right]\right)^{-1} \right) \cdot \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \cdot U_{i,\tau} \right\| = o_p(1) \cdot O_p(1) = o_p(1).$$

The fact that Equation (16) is $o_p(1)$ is a direct consequence of Lemma 1:

$$\begin{aligned}
& \left\| \left(\left(\frac{1}{N} \cdot \sum_{i=1}^{N} X_i \cdot X_i^\intercal\right)^{-1} - \left(E\left[X \cdot X^\intercal\right]\right)^{-1} \right) \cdot \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \cdot \left(\widehat{\mathrm{RIF}}(Y_i;\widehat{q}_\tau) - \mathrm{RIF}(Y_i;q_\tau)\right) \right\| \\
\leq & \left\| \left(\left(\frac{1}{N} \cdot \sum_{i=1}^{N} X_i \cdot X_i^\intercal\right)^{-1} - \left(E\left[X \cdot X^\intercal\right]\right)^{-1} \right) \right\| \cdot \left\| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \cdot \left(\widehat{\mathrm{RIF}}(Y_i;\widehat{q}_\tau) - \mathrm{RIF}(Y_i;q_\tau)\right) \right\| \\
\leq & \sup_y \left| \widehat{\mathrm{RIF}}(y;\widehat{q}_\tau) - \mathrm{RIF}(y;q_\tau) \right| \cdot \left\| \left(\left(\frac{1}{N} \cdot \sum_{i=1}^{N} X_i \cdot X_i^\intercal\right)^{-1} - \left(E\left[X \cdot X^\intercal\right]\right)^{-1} \right) \right\| \cdot \left\| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \right\| \\
= & \; o_p(1) \cdot o_p(1) \cdot O_p(1) = o_p(1).
\end{aligned}$$

Finally, Equation (17) is bounded by the same type of argument:

$$\begin{aligned} & \left\| \left(E\left[X \cdot X^{\intercal}\right]\right)^{-1} \cdot \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \cdot \left(\widehat{\mathrm{RIF}}(Y_i;\widehat{q}_\tau) - \mathrm{RIF}(Y_i;q_\tau)\right) \right\| \\ \leq \ & \sup_y \left| \widehat{\mathrm{RIF}}\left(y;\widehat{q}_\tau\right) - \mathrm{RIF}\left(y;q_\tau\right) \right| \cdot \left\| \left(E\left[X \cdot X^{\intercal}\right]\right)^{-1} \right\| \cdot \left\| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} X_i \right\| \\ = \ & o_p\left(1\right) \cdot O_p\left(1\right) \cdot O_p\left(1\right) = o_p\left(1\right). \end{aligned}$$

■

Proof of Proposition 3: Start by writing $\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{F'}_{Y|X}\left(\widehat{q}_\tau|X=X_i\right) - E\left[F'_{Y|X}\left(q_\tau|X\right)\right]\right)$ as

$$\begin{aligned} & \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{F'}_{Y|X}\left(\widehat{q}_\tau|X=X_i\right) - E\left[F'_{Y|X}\left(q_\tau|X\right)\right]\right) \\ = \ & \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{F'}_{Y|X}\left(\widehat{q}_\tau|X=X_i\right) - \widehat{F'}_{Y|X}\left(q_\tau|X=X_i\right)\right) \\ & + \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{F'}_{Y|X}\left(q_\tau|X=X_i\right) - E\left[F'_{Y|X}\left(q_\tau|X\right)\right]\right) \\ = \ & \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(F'_{Y|X}\left(q_\tau|X=X_i\right) - E\left[F'_{Y|X}\left(q_\tau|X\right)\right]\right) \\ & + \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} l\left(X_i\right) \cdot \left(T_{\tau,i} - p_\tau\left(X_i\right)\right) + o_p\left(1\right) \end{aligned}$$

thus we need to show that $\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{F'}_{Y|X}\left(\widehat{q}_\tau|X=X_i\right) - \widehat{F'}_{Y|X}\left(q_\tau|X=X_i\right)\right)$ is $o_p\left(1\right)$.

$$
\begin{aligned}
& \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{F'}_{Y|X} \left(\widehat{q}_\tau | X = X_i \right) - \widehat{F'}_{Y|X} \left(q_\tau | X = X_i \right) \right) \\
= \; & \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} G_{K(\tau)}(X_i)^\intercal \cdot \left(\widehat{\pi}_{K(\tau)} \cdot \widehat{p}^F_{K(\tau)}(X_i) \cdot \left(1 - \widehat{p}^F_{K(\tau)}(X_i)\right) - \widetilde{\pi}_{K(\tau)} \cdot \widehat{p}^U_{K(\tau)}(X_i) \cdot \left(1 - \widehat{p}^U_{K(\tau)}(X_i)\right)\right) \\
= \; & \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} G_{K(\tau)}(X_i)^\intercal \cdot \left(\widehat{\pi}_{K(\tau)} - \widetilde{\pi}_{K(\tau)} \right) \cdot \widehat{p}^U_{K(\tau)}(X_i) \cdot \left(1 - \widehat{p}^U_{K(\tau)}(X_i)\right) \\
& + \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} G_{K(\tau)}(X_i)^\intercal \cdot \widehat{\pi}_{K(\tau)} \cdot \left(\widehat{p}^F_{K(\tau)}(X_i) \cdot \left(1 - \widehat{p}^F_{K(\tau)}(X_i)\right) - \widehat{p}^U_{K(\tau)}(X_i) \cdot \left(1 - \widehat{p}^U_{K(\tau)}(X_i)\right)\right) \\
\leq \; & \left| \frac{1}{\sqrt{N}} \cdot \sqrt{\left\| \widehat{\pi}_{K(\tau)} - \widetilde{\pi}_{K(\tau)} \right\|^2} \cdot \sup_{x \in \mathcal{X}} \left(\widehat{p}^U_{K(\tau)}(x) \cdot \left(1 - \widehat{p}^U_{K(\tau)}(x)\right) \cdot \sqrt{\left\| G_{K(\tau)}(x) \right\|^2} \right) \right| \\
& + \left| \sqrt{\left\| \widehat{\pi}_{K(\tau)} \right\|^2} \cdot \sup_{x \in \mathcal{X}} \left(\sqrt{\left\| G_{K(\tau)}(x) \right\|^2} \cdot \left(1 - \widehat{p}^F_{K(\tau)}(x) - \widehat{p}^U_{K(\tau)}(x)\right) \right) \right| \cdot \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{p}^F_{K(\tau)}(X_i) - \widehat{p}^U_{K(\tau)}(X
\end{aligned}
$$

By using the fact that $Q_N(\pi)$ and $\widehat{Q}_N(\pi)$ are concave functions in π, we can apply the Convexity Lemma (see, for example, Hjört and Pollard, 1993) which establishes that the minimizers (maximizers) of convex (concave) random functions that can be approximated by the same quadratic function in a neighborhood of the true minimum will be asymptotically equivalent. Thus, given that as N increases $\widehat{q}_\tau$ gets closer in probability to q_τ, $\widehat{Q}_N(\cdot)$ will get closer uniformly in probability to $Q_N(\cdot)$ and one can show that $\left\| \widehat{\pi}_{K(\tau)} - \widetilde{\pi}_{K(\tau)} \right\| = o_p(1)$.

Also, according to the assumptions of Hirano, Imbens and Ridder (2003)

$$
\sup_{x \in \mathcal{X}} \left(\widehat{p}^U_{K(\tau)}(x) \cdot \left(1 - \widehat{p}^U_{K(\tau)}(x)\right) \cdot \sqrt{\left\| G_{K(\tau)}(x) \right\|^2} \right) = O_p(1)
$$

and therefore,

$$
\left| \frac{1}{\sqrt{N}} \cdot \sqrt{\left\| \widehat{\pi}_{K(\tau)} - \widetilde{\pi}_{K(\tau)} \right\|^2} \cdot \sup_{x \in \mathcal{X}} \left(\widehat{p}^U_{K(\tau)}(x) \cdot \left(1 - \widehat{p}^U_{K(\tau)}(x)\right) \cdot \sqrt{\left\| G_{K(\tau)}(x) \right\|^2} \right) \right| = o_p(1)
$$

Finally, by the assumptions involving the series approximation,

$$
\sqrt{\left\| \widehat{\pi}_{K(\tau)} \right\|^2} \cdot \sup_{x \in \mathcal{X}} \left(\sqrt{\left\| G_{K(\tau)}(x) \right\|^2} \cdot \left(1 - \widehat{p}^F_{K(\tau)}(x) - \widehat{p}^U_{K(\tau)}(x)\right) \right) = O_p(1)
$$

and

$$
\begin{aligned}
& \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{p}^{F}_{K(\tau)}(X_i) - \widehat{p}^{U}_{K(\tau)}(X_i) \right) \right| \\
= & \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} L(H_{K(\tau)}(X_i)^{\intercal} \widehat{\pi}_{K(\tau)}) - L(H_{K(\tau)}(X_i)^{\intercal} \widetilde{\pi}_{K(\tau)}) \right| \\
\leq & \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} L'(H_{K(\tau)}(X_i)^{\intercal} \widehat{\pi}^{*}_{K(\tau)}) \right| \cdot \sqrt{\left\| \widehat{\pi}_{K(\tau)} - \widetilde{\pi}_{K(\tau)} \right\|^2} \\
= & O_p(1) \cdot o_p(1) = o_p(1)
\end{aligned}
$$

where $\widehat{\pi}^{*}_{K(\tau)}$ is an intermediate value between $\widehat{\pi}_{K(\tau)}$ and $\widetilde{\pi}_{K(\tau)}$. Thus, we have that

$$
\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{F'}_{Y|X}(\widehat{q}_{\tau}|X = X_i) - \widehat{F'}_{Y|X}(q_{\tau}|X = X_i) \right) = o_p(1)
$$

■

References

David H. Autor, D.H., Katz, L.F. and M.S. Kearney (2006): "The Polarization of the U.S. Labor Market," *American Economic Review* 96, 189–194.

Buchinsky, M. (1994): "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, 62, 405-58.

Card, D. (1996): "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64, 957-79.

Chamberlain, G. (1994): "Quantile Regression Censoring and the Structure of Wages," in *Advances in Econometrics* ed. by C. Sims. Elsevier: New York.

Deaton, A. and S. Ng (1998): "Parametric and Nonparametric Approaches to Price and Tax Reform," *Journal fo the American Statistical Association*, 93, 900–909.

DiNardo, J. and T. Lemieux (1997): "Diverging Male Wage Inequality in the United States and Canada, 1981-1988: Do Institutions Explain the Difference?" *Industrial & Labor Relations Review*, 50, 629–651.

FERNHOLZ, L.T. (2001): "On Multivariate Higher Order von Mises Expansions," *Metrika*, 53, 123–140.

HJORT, N.L. AND D. POLLARD (1993): "Asymptotics for minimisers of convex processes," Statistical Research Report, Yale University.

KOENKER, R. AND G. BASSETT (1978): "Regression Quantiles", *Econometrica*, 46, 33–50.

KOENKER, R. AND K.F. HALLOCK (2001): "Quantile Regression," *Journal of Economic Perspectives*, 15, 143-56.

HAMPEL, F.R. (1968): "Contribution to the Theory of Robust Estimation," Ph.D. Thesis, University of California at Berkeley.

HAMPEL, F.R. (1974): "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 60, 383–393.

HAMPEL, F.R., E.M. RONCHETTI, P.J. ROUSSEEUW, AND W.A. STAHEL (1986): *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley.

HARDLE, W. AND STOKER, T. M. (1989) "Investigating Smooth Multiple Regression by the Method of Average Derivatives", *Journal of the American Statistical Association,* 84, 986–95.

HIRANO, K., IMBENS, G. W. AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189.

LEMIEUX, T. (2006a): "Postsecondary Education and Increasing Wage Inequality," *American Economic Review* 96, 195-199.

LEMIEUX, T. (2006b): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?, *American Economic Review* 96, forthcoming June.

LEWBEL, A. (2004) "Simple Estimators for Hard Problems: Endogeneity in Discrete Choice Related Models, ' mimeo.

LEWBEL, A. (2005) "Simple Endogenous Binary Choice and Selection Panel Model Estimators,' mimeo.

von Mises, R. (1947): "On the asymptotic distribution of differentiable statistical functions". *The Annals of Mathematical Statistics*, 18, 309—348.

Newey, W. K. and T.M. Stoker (1993): "Efficiency of Weighted Average Derivative Estimators and Index Models," *Econometrica*, 61, 1199–1223.

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

Withers, C.S. (1983). "Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals," *The Annals of Mathematical Statistics*, 11, 577—587.

Wooldridge, J. M. (2004). "Estimating Average Partial Effects Under Conditional Moment Independence Assumptions," mimeo. Michigan State University.

Table 1: Comparing OLS, Conditional quantile regressions (CQR) and unconditional quantile regressions (UQR), 1983-85 CPS data from men.

	OLS	10th centile		50th centile		90th centile	
		UQR	CQR	UQR	CQR	UQR	CQR
Union coverage	0.179	0.198	0.288	0.349	0.195	-0.137	0.088
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)	(0.004)
Non-white	-0.134	-0.118	-0.139	-0.169	-0.134	-0.101	-0.120
	(0.003)	(0.005)	(0.004)	(0.004)	(0.003)	(0.005)	(0.005)
Married	0.140	0.197	0.166	0.162	0.146	0.044	0.089
	(0.002)	(0.003)	(0.003)	(0.004)	(0.002)	(0.004)	(0.004)
Education							
Elementary	-0.351	-0.311	-0.279	-0.469	-0.374	-0.244	-0.357
	(0.004)	(0.008)	(0.006)	(0.006)	(0.004)	(0.005)	(0.007)
HS Dropout	-0.190	-0.349	-0.127	-0.202	-0.205	-0.069	-0.227
	(0.003)	(0.006)	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)
Some college	0.133	0.059	0.058	0.185	0.133	0.156	0.172
	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)
College	0.406	0.199	0.252	0.481	0.414	0.592	0.548
	(0.003)	(0.004)	(0.004)	(0.005)	(0.003)	(0.008)	(0.005)
Post-graduate	0.478	0.140	0.287	0.541	0.482	0.859	0.668
	(0.004)	(0.004)	(0.004)	(0.005)	(0.003)	(0.010)	(0.005)
Experience							
0-4	-0.545	-0.599	-0.333	-0.641	-0.596	-0.454	-0.650

	(0.004)	(0.007)	(0.005)	(0.006)	(0.004)	(0.008)	(0.007)
5-9	-0.267	-0.082	-0.191	-0.360	-0.279	-0.377	-0.319
	(0.004)	(0.005)	(0.005)	(0.006)	(0.004)	(0.008)	(0.006)
10-14	-0.149	-0.040	-0.098	-0.185	-0.152	-0.257	-0.188
	(0.004)	(0.004)	(0.005)	(0.006)	(0.004)	(0.009)	(0.006)
15-19	-0.056	-0.024	-0.031	-0.069	-0.060	-0.094	-0.077
	(0.004)	(0.004)	(0.005)	(0.006)	(0.004)	(0.009)	(0.007)
25-29	0.028	0.001	0.001	0.034	0.029	0.063	0.038
	(0.004)	(0.005)	(0.006)	(0.007)	(0.005)	(0.011)	(0.007)
30-34	0.034	0.004	-0.007	0.038	0.033	0.063	0.064
	(0.004)	(0.005)	(0.006)	(0.007)	(0.005)	(0.011)	(0.008)
35-39	0.042	0.021	-0.014	0.041	0.043	0.073	0.095
	(0.005)	(0.005)	(0.006)	(0.007)	(0.005)	(0.011)	(0.008)
40+	0.005	0.042	-0.066	0.002	0.017	-0.030	0.061
	(0.005)	(0.006)	(0.007)	(0.007)	(0.005)	(0.010)	(0.008)
Constant	1.742	0.970	1.145	1.732	1.744	2.512	2.332
	(0.004)	(0.005)	(0.005)	(0.006)	(0.004)	(0.008)	(0.006)

Figure 1a. Unconditional and Conditional Quantile Regressions Estimates for the Birthweight Model

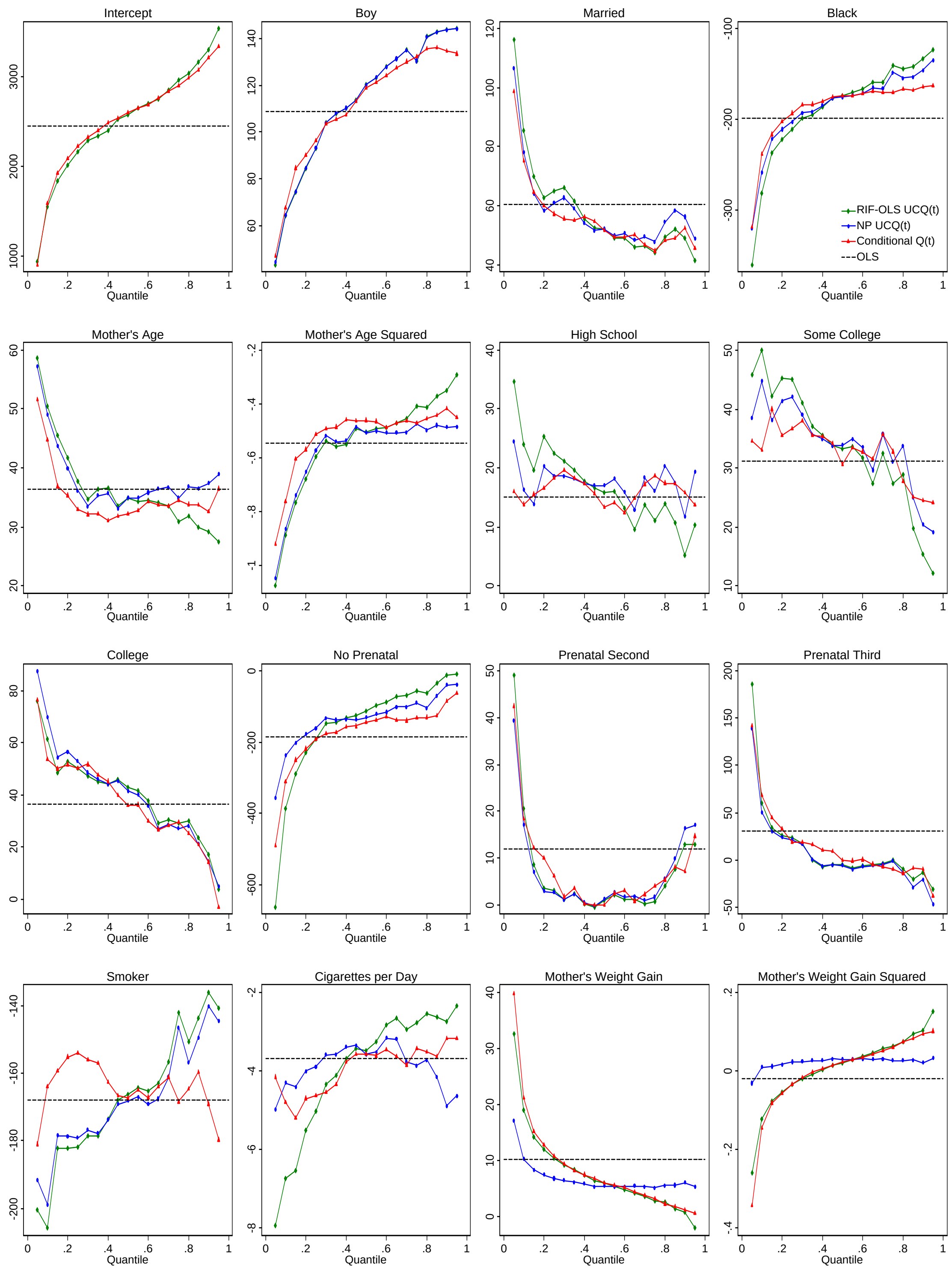


Figure 1b. Significance of the Differences between Different Quantile Estimates

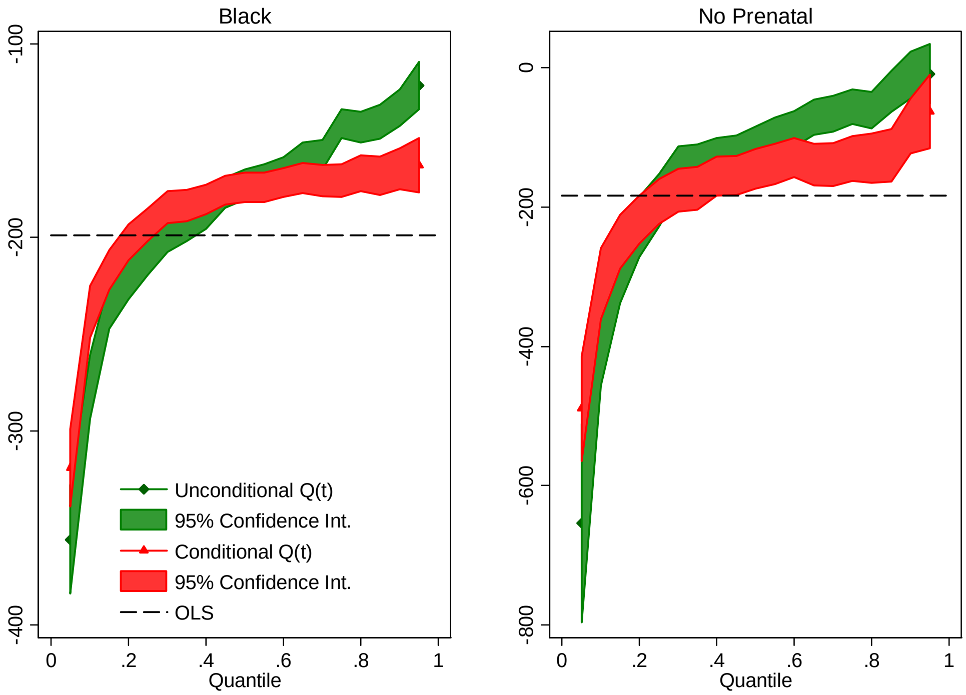


Figure 2a. Unconditional and Conditional Quantile Estimates for the Log Wages Model, Men 1983-1985

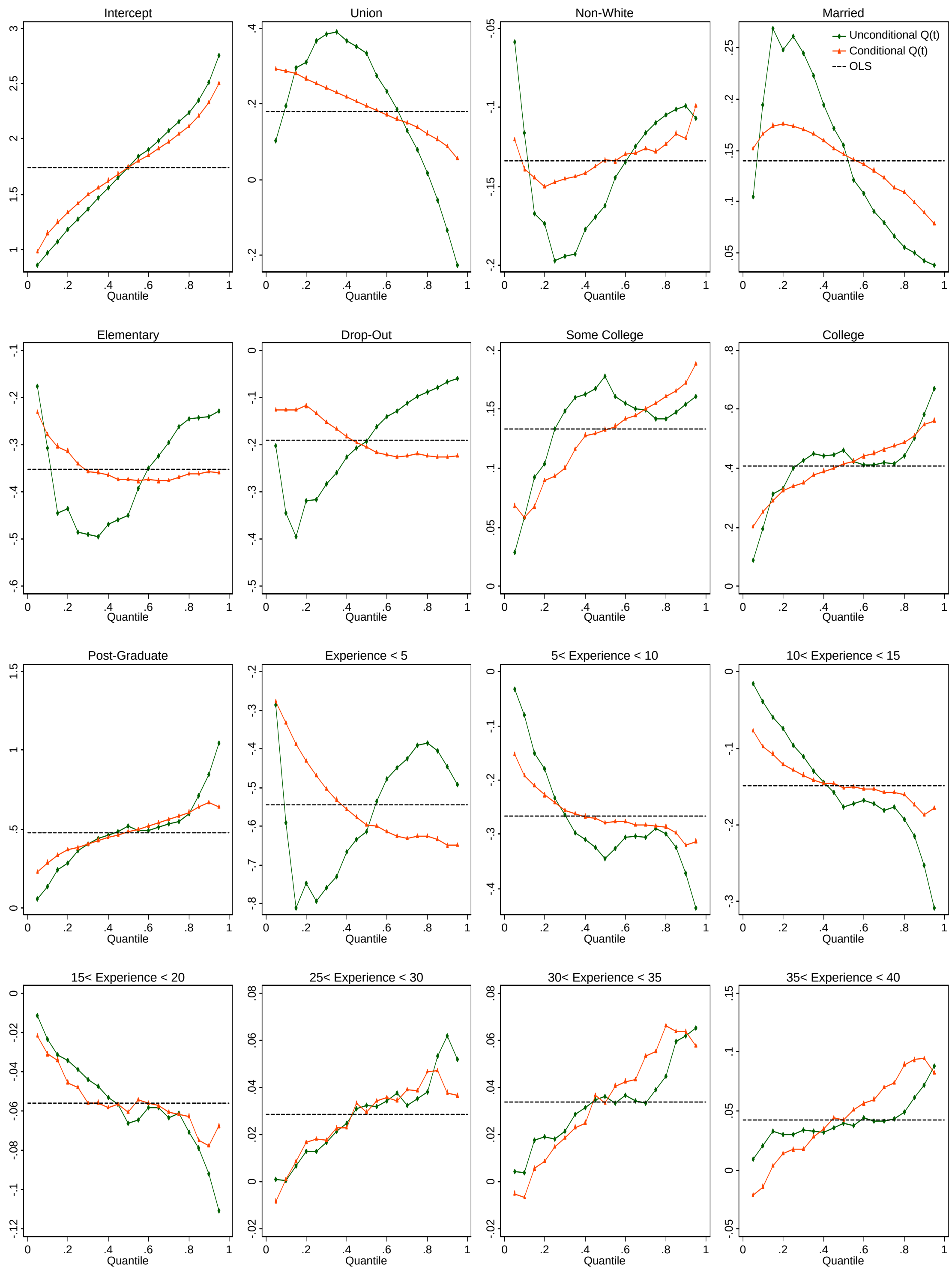


Figure 2b. Unconditional and Conditional Quantile Estimates for the Log Wages Model, Men 2003-2005

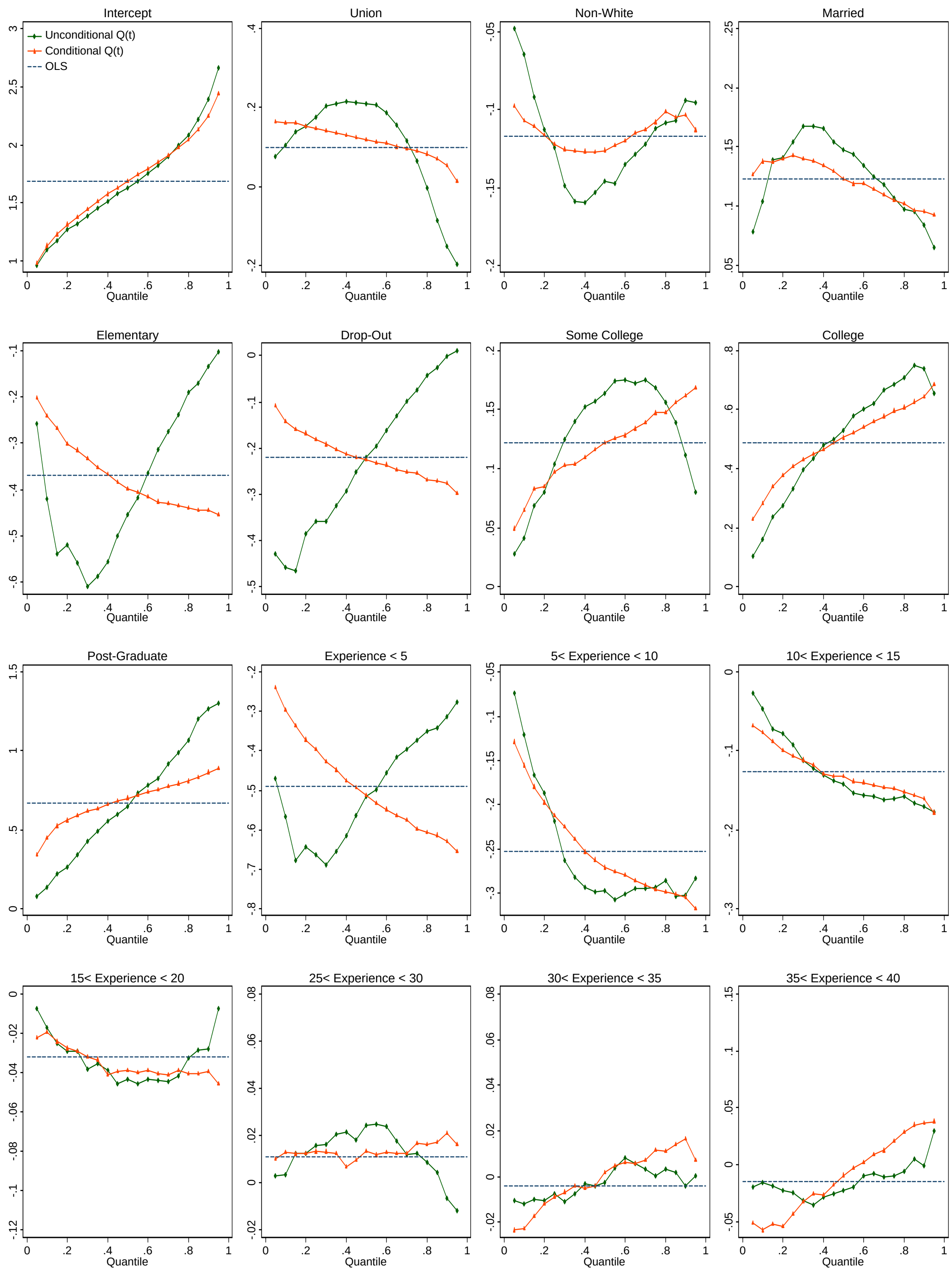


Figure 2c. RIF-OLS and Non-Parametric Unconditional Quantile Regressions Estimates for the Log Wages Model, Men 1983-1985

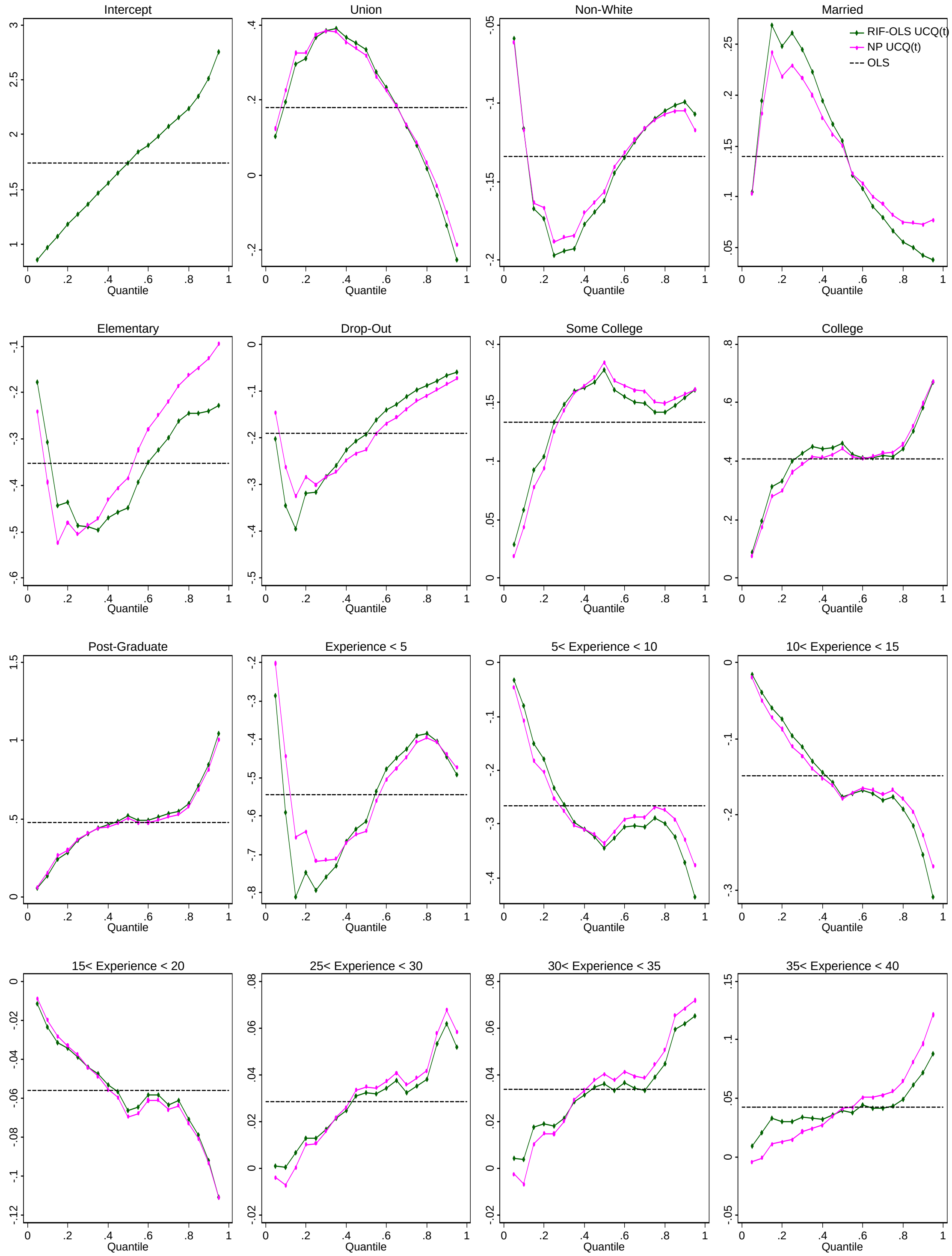


Figure 2d. Sensitivity of Unconditional Estimates for the Log Wages Model, Men 1983-1985

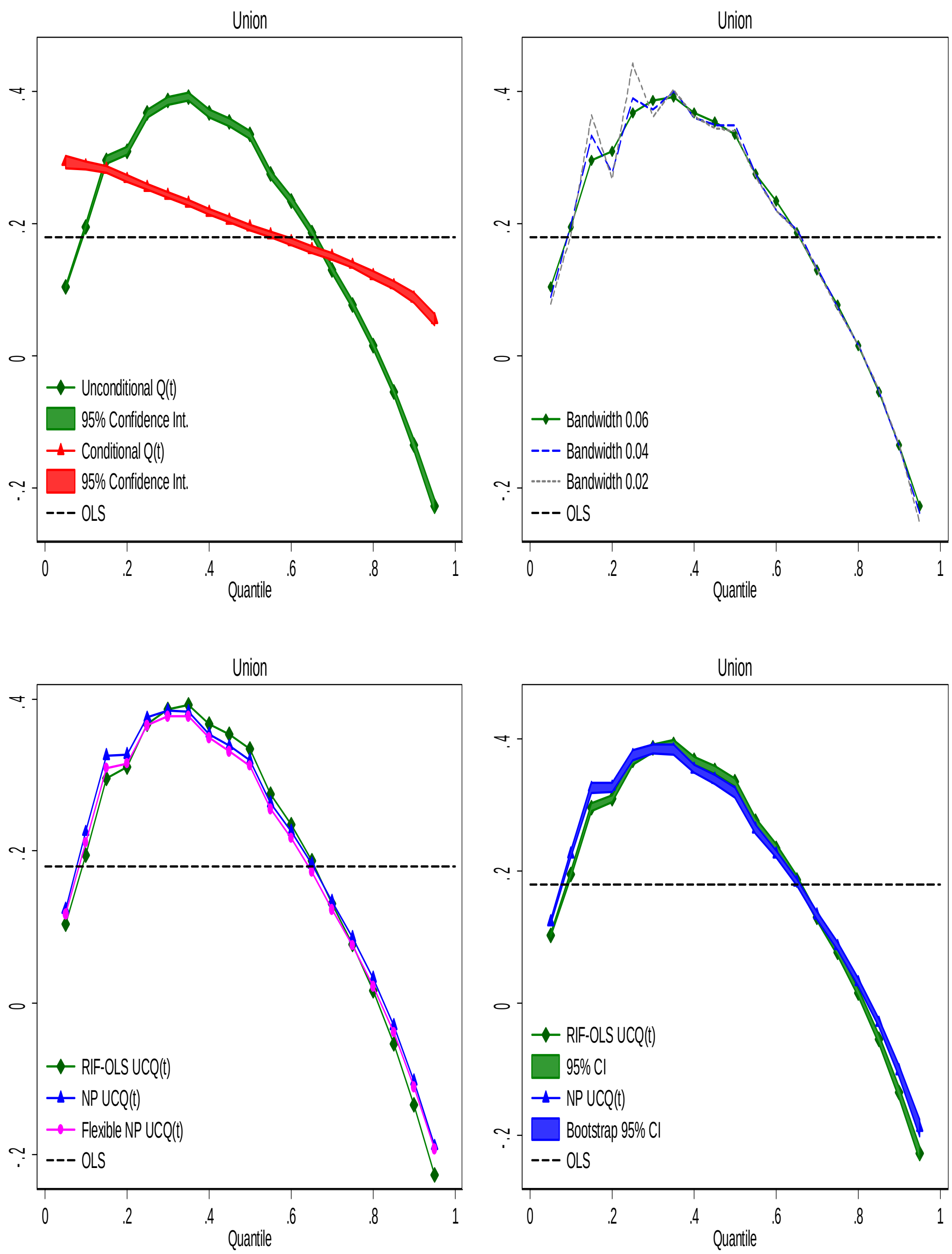


Figure 3: Approximation error (relative to reweighting) when predicting the effect of changes in the unionization rate using the unconditional quantile regression

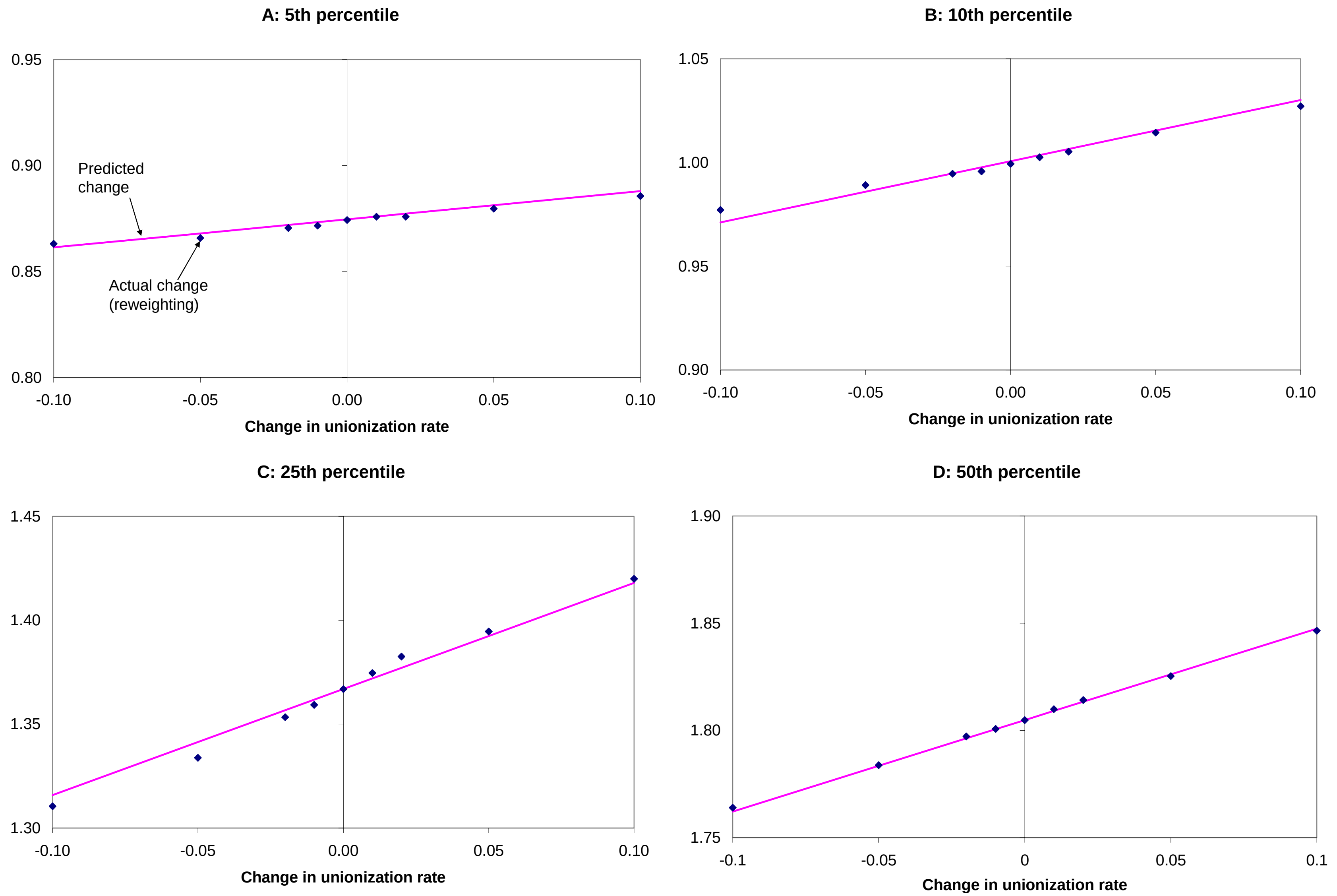


Figure 3: Continuation

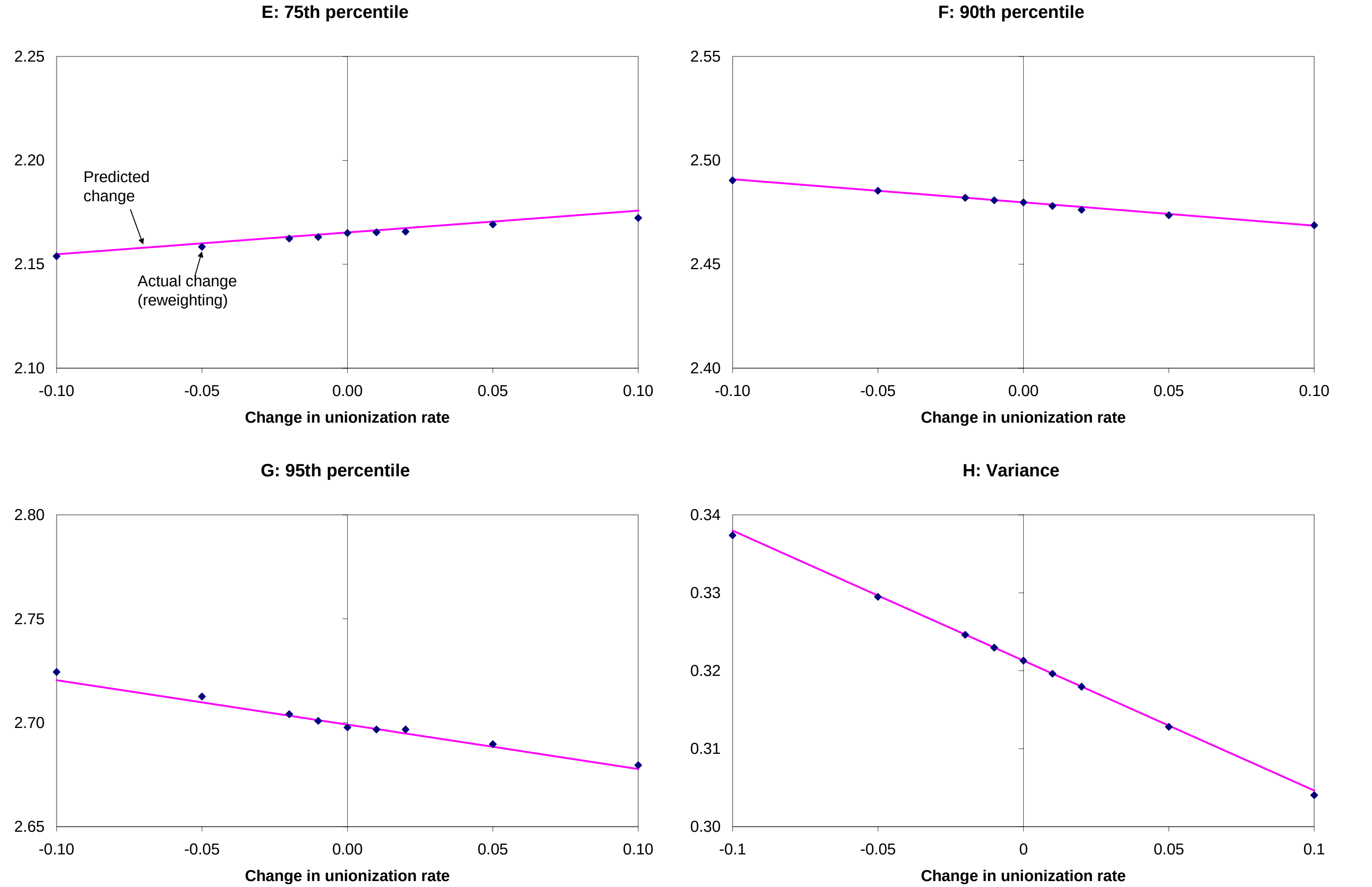


Figure A1. Probability Density Functions of Birthweight

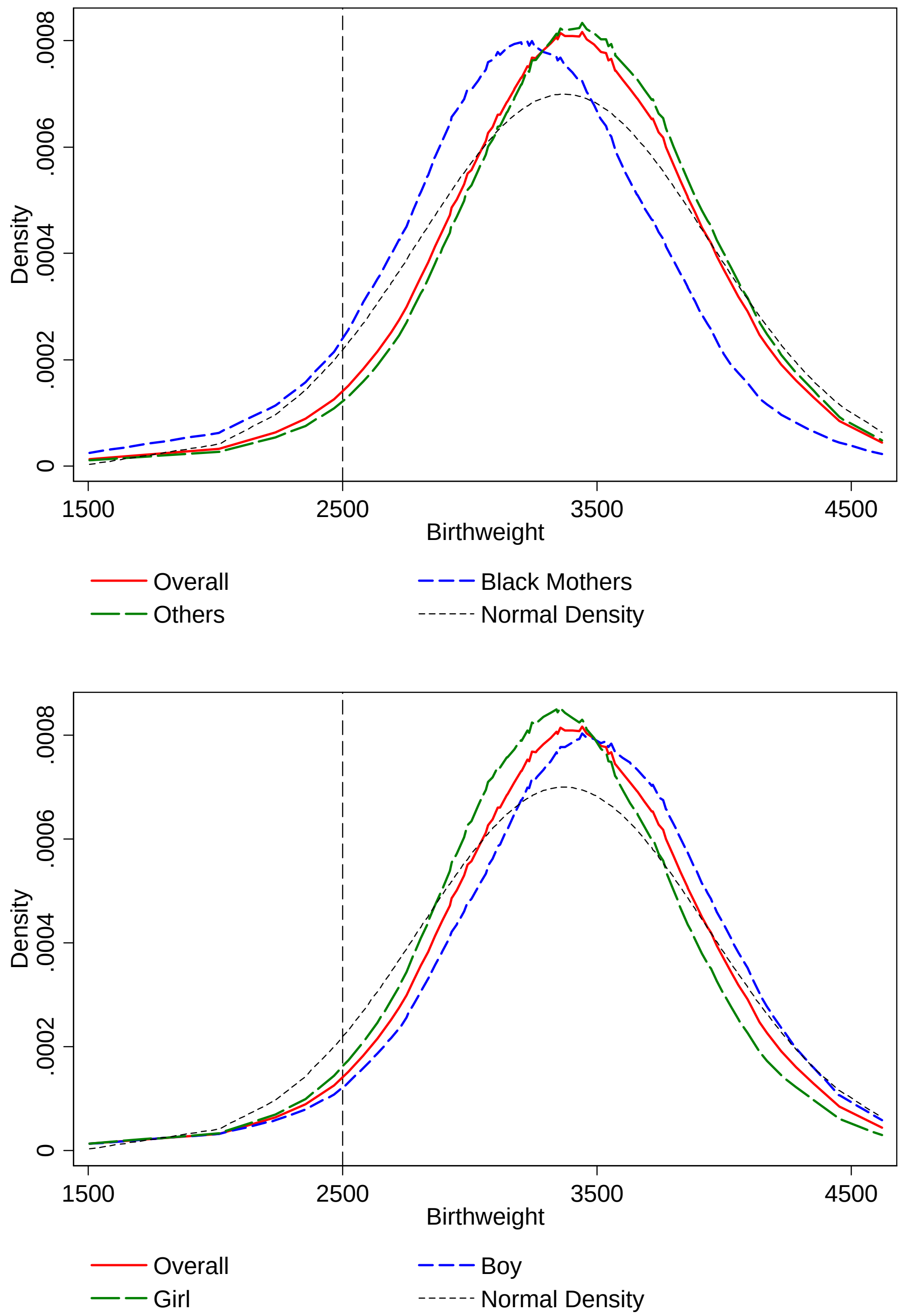
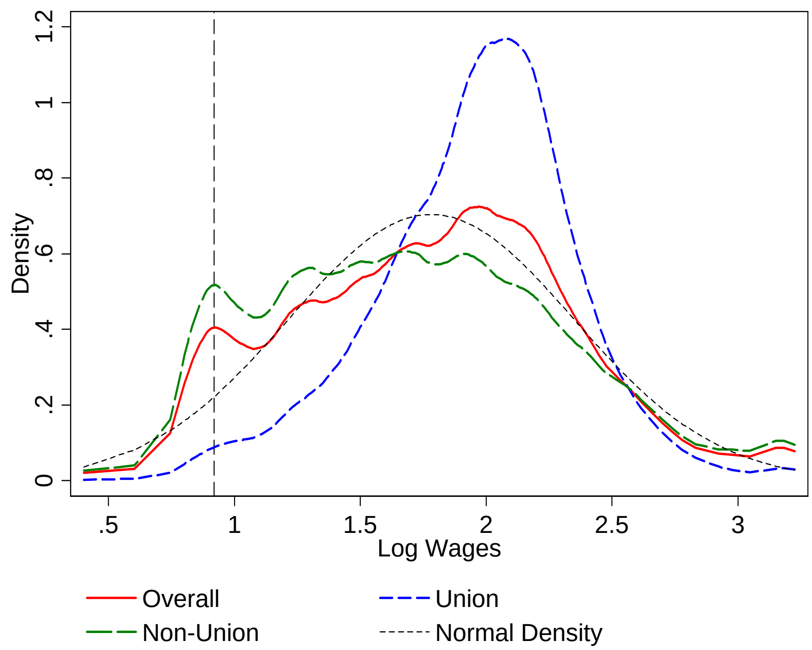


Figure A2. Probability Density Functions of Log Wages,
a) Men 1983-85



b) Men 2003-05