Unconditional Quantile Regressions*

Sergio Firpo,

Pontifícia Universidade Católica -Rio

Nicole M. Fortin, and Thomas Lemieux

University of British Columbia

October 2006
Comments welcome

Abstract

We propose a new regression method for modelling unconditional quantiles of an outcome variable as a function of the explanatory variables. The method consists of running a regression of the (recentered) influence function of the unconditional quantile of the outcome 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 changes in explanatory variables on a given unconditional quantile, just like the regression coefficients are used in the case of the mean. We also discuss how our approach can be generalized to other distributional statistics besides quantiles.

Keywords: *Influence Functions, Unconditional Quantile, Quantile Regressions.*

*We are indebted to Richard Blundell, David Card, Vinicius Carrasco, Marcelo Fernandes, Chuan Goh, Joel Horowitz, Shakeeb Khan, Roger Koenker, Thierry Magnac, Whitney Newey, Geert Ridder, Jean-Marc Robin, Hal White and seminar participants at the CESG 2005 Meeting, CAEN-UFC and Econometrics in Rio 2006 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, and SSHRC for financial support.

1 Introduction

One important reason for the popularity of OLS regressions in economics is that they provide consistent estimates of the impact of an explanatory variable, X, on the *population unconditional mean* of an outcome variable, Y. This important property stems from the fact that the conditional mean, $E[Y|X]$, averages up to the unconditional mean, $E[Y]$, thanks to the law of iterated expectations. As a result, a linear model for conditional means, $E[Y|X] = X\beta$, implies that $E[Y] = E[X]\beta$, and OLS estimates of β also indicate what is the impact of X on the population average of Y. Many important applications of regression analysis crucially rely on this important property. For example, consider the effect of education on earnings. Oaxaca-Blinder decompositions of the earnings gap between blacks and whites or men and women, growth accounting exercises (contribution of education to economic growth), and policy intervention analyses (average treatment effect of education on earnings) all crucially depend on the fact that OLS estimates of β also provide an estimate of the effect of increasing education on the average earnings in a given population.

When the underlying question of economic and policy interest concerns other aspects of the distribution of Y, however, estimation methods that "go beyond the mean" have to be used. A convenient way of characterizing the distribution of Y is by computing its quantiles.[1] Indeed, quantile regressions have become an increasingly popular method for addressing distributional questions as they provide a convenient way of modelling any conditional quantile of Y as a function of X. Unlike conditional means, however, conditional quantiles do not average up to their unconditional population counterparts. As a result, the estimates obtained by running a quantile regression cannot be used to estimate the impact of X on the corresponding unconditional quantile. This means that existing methods cannot be used to answer a question as simple as "what is the effect on median earnings of increasing everybody's education by one year, holding everything else constant?".

In this paper, we propose a new computationally simple regression method that can be used to estimate the effect of explanatory variables on the unconditional quantiles of the outcome variable. To contrast our approach from commonly used *conditional* quantile regressions (Koenker and Bassett, 1978), we call our regression method *unconditional*

[1] Discretized versions of the distribution functions can be calculated using quantiles, as well many inequality measurements such as, for instance, quantile ratios, inter-quantile ranges, concentration functions, and the Gini coefficient. This suggests modelling quantiles as a function of the covariates to see how the whole distribution of Y responds to changes in the covariates.

quantile regressions. Our approach builds upon the concept of the influence function (IF), a widely used tool in robust estimation of statistical or econometric models. The IF represents, as the name suggests, the influence of an individual observation on the distributional statistic. The influence functions of commonly used statistics are very well known. For example, the influence function for the mean $\mu = E[Y]$ is simply the demeaned value of the outcome variable, $Y - \mu$.[2]

Adding back the statistic to the influence function yields what we call the *Recentered Influence Function* (RIF), which is simply Y in the case of the mean. More generally, the RIF can be thought of as the contribution of an individual distribution to a given distributional statistic. It is easy to compute the RIF for quantiles or most distributional statistics. For a quantile, the RIF is simply $q_\tau + \mathrm{IF}(Y, q_\tau)$, where $q_\tau = Q_\tau[Y]$ is the population τ-quantile of the unconditional distribution of Y, and $\mathrm{IF}(Y, q_\tau)$ is its influence function, known to be equal to $(\tau - \mathbb{I}\{Y \leq q_\tau\}) \cdot f_Y^{-1}(q_\tau)$; $\mathbb{I}\{\cdot\}$ is an indicator function, and $f_Y(\cdot)$ is the density of the marginal distribution of Y.[3]

We propose, in the simplest case, to estimate how the covariates X affect the statistic (functional) ν of the unconditional distribution of Y we by running standard regressions of the RIF on the explanatory variables. This regression what we call the RIF-Regression Model, $E[\mathrm{RIF}(Y, \nu) | X]$. Since the RIF for the mean is just the value of the outcome variable, Y, a regression of $\mathrm{RIF}(Y, \mu)$ on X is the same as standard regression of Y on X. In our framework, this is why OLS estimates are valid estimates of the effect of X on the unconditional mean of Y. More importantly, we show that this property extends to any other distributional statistic. For the τ-quantile, we show the conditions under which a regression of $\mathrm{RIF}(Y, q_\tau)$ on X can be used to consistnetly estimate the effect of X on the unconditional τ-quantile of Y. In the case of quantiles, we call the RIF-regression model, $E[\mathrm{RIF}(Y, q_\tau) | X]$, an *Unconditional Quantile Regression.* We will define in Section 4 the exact population parameters that we estimate with this regression, that is, we provide a more formal definition of what we mean by the "effect" of X on the unconditional τ-quantile of Y.

We view our method as a very important complement to conditional quantile regressions. Of course, in some settings quantile regressions are the appropriate method to use.[4] For instance, quantile regressions are a useful descriptive tool that provide a par-

[2]Observations with value of Y far away from the mean have more influence on the mean than observations with value of Y close to the mean.

[3]We define the unconditional quantile operator as $Q_\tau[\cdot] \equiv \inf_q \Pr[\cdot \leq q] \geq \tau$. Similarly, the conditional (on $X = x$) quantile operator is defined as $Q_\tau[\cdot | X = x] \equiv \inf_q \Pr[\cdot \leq q | X = x] \geq \tau$.

[4]See, for example, Buchinski (1994) and Chamberlain (1994) for applications of conditional quantile

simonious representation of the conditional quantiles. Unlike standard OLS regression estimates, however, quantile regression estimates cannot be used to assess the more general economic or policy impact of a change of X on the corresponding quantile of the overall (or unconditional) distribution of Y. While OLS estimates can either be used as estimates of the effect of X on the conditional or the unconditional mean, one has to be much more careful in deciding what is the ultimate object of interest in the case of quantiles.

For instance, consider of the effect of union status on log wages (one of the example of quantile regression studied by Chamberlain, 1994). If the OLS estimate of the effect of union on log wages is 0.2, this means that a decline of 1 percent in the rate of unionization will lower average wages by 0.2 percent. But if the estimated effect of unions (using quantile regressions) on the *conditional* 90^{th} quantile is 0.1, this does not mean that a decline of 1 percent in the rate of unionization would lower the *unconditional* 90^{th} quantile by 0.1 percent. In fact, we show in an empirical application in Section 6 that unions have a positive effect on the *conditional* 90^{th} quantile, but a negative effect on the *unconditional* 90^{th} quantile. If one is interested in the overall effect of unions on wage inequality, our unconditional quantile regressions should be used to obtain the effect of unions at different quantiles of the unconditional distribution. Using quantile regressions to estimate the overall effect of unions on wage inequality would yield a misleading answer in this particular case.

The structure of the paper is as follows. In the next section (Section 2), we present the basic model and define two key objects of interest in the estimation: the "policy effect" and the "unconditional partial effect". In Section 3, we establish the general concepts of recentered influence functions. 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 show how unconditional quantile regressions can be used to estimate either the policy effect and the unconditional (quantile) partial effect. We discuss the general case and specific examples with an explicit structural model linking Y and X. We discuss estimation issues in Section 5. Section 6 presents two applications of our method: the impact of unions on the distribution of wages, and the determinants of the distribution of infants' birthweight (as in Koenker and Hallock, 2001). We conclude in Section 7.

---

regressions to various issues related to the wage structure.

2 Model Setup and Parameters of Interest

Before presenting the estimation method, it is important to clarify exactly what the unconditional quantile regressions seek to estimate. 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. Assume that the dependent variable Y is a function of observables X and unobservables ε, according to the following model:

$$Y = h(X,\varepsilon),$$

where $h(\cdot,\cdot)$ is an unknown mapping, assumed to be monotonic in ε. Note that using a flexible function $h(\cdot,\cdot)$ is important for allowing for rich distributional effect of X on Y.[5]

We are primarily interested in estimating two population parameters, the unconditional quantile partial effect and the policy effect, using unconditional quantile regressions. We now formally define these two parameters.

Unconditional Quantile Partial Effect (UQPE)

By analogy with a standard regression coefficient, our first object of interest is the effect on an unconditional quantile of a small increase t in the explanatory variable X. This effect of a small change in X on the τ^{th} quantile of the unconditional distribution of Y, is defined as:

$$\alpha(\tau) = \lim_{t\downarrow 0}\frac{Q_\tau\left[h\left(X+t,\varepsilon\right)\right]-Q_\tau\left[Y\right]}{t}$$

where $Q_\tau\left[Y\right]$ is the τ^{th} quantile of the unconditional distribution of the random variable Y.[6]

We call this parameter, $\alpha(\tau)$, the *unconditional quantile partial effect* (UQPE), by analogy with Wooldridge (2004) unconditional average partial effect (UAPE), which is defined as $E\left[\partial E\left[Y|X\right]/\partial x\right]$.[7]The link between the UQPE and Wooldridge's UAPE can

[5]A number of recent studies also use general nonseparable models to investingate a number of related issues. See, for example, Chesher, 2003, Florens et al., 2003, and Imbens and Newey, 2005.

[6]Note that to simplify the exposition we are treating X as univariate. However, this is easily generalized to the multivariate case by defining for each $j=1,...,k$,

$$\alpha_j(\tau) = \lim_{t_j\downarrow 0}\frac{Q_\tau\left[h\left(\left[X_j+t_j;X_{-j}\right],\varepsilon\right)\right]-Q_\tau\left[Y\right]}{t_j}$$

where $X=[X_j+t_j;X_{-j}]$.

[7]The UAPE can also be defined in the same way as the UQPE: $UAPE=\lim_{t\downarrow 0}\frac{E[h(X+t,\varepsilon)]-E[Y]}{t}$

be established using a result of Section 3 where we show that for any statistic ν defined as a functional of the unconditional distribution of Y, $\alpha(\nu)$ is exactly equal to $E\left[\partial E\left[\mathrm{RIF}(Y,\nu)|X\right]/\partial x\right]$. Given that $\mathrm{RIF}(Y,\mu)=Y$, we have that $\alpha(\mu)$ is indeed UAPE. This is why our parameter of interest, $\alpha(\tau)$, which is equal $E\left[\partial E\left[\mathrm{RIF}(Y,q_\tau)|X\right]/\partial x\right]$ (see Sections 3 and 4), is called the UQPE.

Similarly, by analogy with Wooldridge's (2004) conditional average partial effect (CAPE) defined as $\partial E\left[Y|X\right]/\partial x$, we can think of conditional quantile regressions as a method for estimating the partial effects of the conditional quantiles of Y given a particular value X. We refer to this type of quantile partial effects as "conditional quantile partial effects" (CQPE) and define them as $\partial Q_\tau\left[Y|X\right]/\partial x=\lim_{t\downarrow 0}\frac{Q_\tau[h(X+t,\varepsilon)|X]-Q_\tau[Y|X]}{t}$ (see Section 4 for more detail).

Note that while UAPE equals the average CAPE, the same relationship does not hold between the UQPE and the CQPE. We will indeed show in Section 4 that the $\alpha(\tau)$ is equal to a complicated weighted average of quantile regression coefficients over the whole range of conditional quantiles (i.e. for τ going from 0 to 1).

Policy Effect

We are also interested in estimating the impact of a more general change in X on the τ^{th} quantile of Y. Consider the "intervention" or "policy change" proposed by Stock (1989) and Imbens and Newey (2005), where X is replaced by the function $\ell(X)$, $\ell:\mathcal{X}\rightarrow\mathcal{X}$. For example, if X represents years of schooling, a compulsory high school completion program aimed at making sure everyone completes grade twelve would be captured by the policy function $\ell(\cdot)$, where $\ell(X)=12$ if $x\leqq 12$, and $\ell(X)=x$ otherwise. We define $\delta_\ell(\tau)$ as the effect of the policy on the τ^{th} quantile of Y, where

$$\delta_\ell(\tau)=Q_\tau\left[h\left(\ell\left(X\right),\varepsilon\right)\right]-Q_\tau\left[Y\right].$$

In the case of the mean we have

$$\delta_\ell(\mu)=E\left[h\left(\ell\left(X\right),\varepsilon\right)\right]-E\left[Y\right]=E\left[E\left[h\left(\ell\left(X\right),\varepsilon\right)-h\left(X,\varepsilon\right)|X\right]\right]$$

which is simply the mean of the policy effect proposed by Stock (1989).

The main contribution of the paper is to show that, as in the case of the mean, a regression framework where the outcome variable Y is replaced by $\mathrm{RIF}(Y,q_\tau)$ can be used to estimate the unconditional partial effect $\alpha(\tau)$ and the policy effect $\delta_\ell(\tau)$ for quantiles. We show this formally after introducing some general concepts in Section 3. Since these general concepts hold for any functional of the distribution of interest, the proposed regression framework extends to other distributional statistics such as the variance or the Gini coefficient.

Before introducing theses general concepts, a few remarks or caveats are in order. First, both the UQPE and the policy effect involve manipulations of the explanatory variables that can also be modelled as changes in the distribution of X, $F_X(x)$. Under the "policy change", $\ell(X)$, the resulting counterfactual distribution is given by $G_X(x) = F_X(\ell^{-1}(x))$.[8] Representing manipulations of X in terms of the counterfactual distribution, $G_X(x)$, makes it easier to derive the impact of the manipulation on $F_Y(y)$, the unconditional distribution of the outcome variable Y. By definition, the unconditional distribution function of Y can be written as

$$F_Y(y) = \int F_{Y|X}(y|X=x) \cdot dF_X(x). \tag{1}$$

Under the assumption that the conditional distribution $F_{Y|X}(\cdot)$ is unaffected by manipulations of the distribution of X, a counterfactual distribution of Y, G_Y, can be obtained by replacing $F_X(x)$ with $G_X(x)$:

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

Although the construction of this counterfactual distribution looks purely mechanical, important economic conditions are imbedded in the assumption that $F_{Y|X}(\cdot)$ is unaffected by manipulations of X. Because $Y = h(X, \varepsilon)$, a sufficient condition for $F_{Y|X}(\cdot)$ to be unaffected by manipulations of X is that ε is independent of X. For the sake of simplicity, we implicitly maintain this assumption throughout the paper, although it may be too strong in specific cases.[9] Since $h(X, \varepsilon)$ can be very flexible, independence of ε and X still allow for unobservables to have rich distributional impacts.

[8] If $\ell(\cdot)$ is not globally invertible we may actually break down the support of X in regions where the function is invertible. This allows $G_X(x) = F_X\left(\ell^{-1}(x)\right)$ to be very general.

[9] The independence assumption can easily be relaxed. For instance, if $X = (X_1, X_2)$ and only X_1 is being manipulated, it is sufficient to assume that ε is independent of X_1 conditional on X_2. This conditional independence assumption is similar to the "ignorability" or "unconfoundedness" assumption commonly used in the program evaluation literature. Independence between ε and of X could also be obtained by conditioning on a control function constructed using instrumental variables, as in Chesher (2003), Florens et al., (2003), and Imbens and Newey (2005).

A second remark is that, as in the case of a standard regression model for conditional means, there is no particular reasons to expect RIF regression models to be linear in X. In fact, in the case of quantiles we show in Section 4 that even for the most basic linear model, $h(X, \varepsilon) = X\beta + \varepsilon$, the RIF regression is not linear. Fortunately, the non-linear nature of the RIF regression is closely related to the problem of estimating a regression model for a dichotomous dependent variable. Widely available estimation procedures (probit or logit) can thus used to deal the special nature of the dependent variable (RIF for quantiles). In the empirical section (Section 6) we show that, in practice, different regression methods yield very similar estimates of the UQPE. This finding is not surprising in light of the "common empirical wisdom" that probits, logits, and linear probability models all yield very similar estimates of marginal effects in a wide variety of cases.

A final remark is that while our regression method yields exact estimates of the UQPE, it only yields a first order approximation of the policy effect $\delta_\ell(\tau)$. In other words, it is not clear how accurate our estimates of $\delta_\ell(\tau)$ would be in the case of large interventions that involve very substantial changes in the distribution of X. How good the approximation is turns out to be an empirical question. We show in the case of unions and wage inequality (Section 6) that our method yields very accurate results even in case of economically large changes in the rate of unionization.

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 is central 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. The derivations are developed for general functionals of the distribution; they will be applied to quantiles (and the mean) in the next section.

3.1 Definition and Properties of Recentered Influence Functions

We begin by recalling the theoretical foundation of the definition of the influence functions following Hampel et al. (1986). For notational simplicity, in this subsection we drop the subscript Y on F_Y and G_Y. 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 $|\nu(F)| < +\infty$. In our setting, F is the CDF of the outcome variable Y, while $\nu(F)$ is a distributional statistic such as a quantile. Following Huber (1977), we say that $\nu(\cdot)$ is Gâteaux differentiable 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 $F_{t,G}$

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

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 $d(G-F)(y)$:

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

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

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

By a normalization argument, $\mathrm{IF}(y;\nu,F)$, the influence function of ν evaluated at y and at the starting distribution F will be written as $\mathrm{IF}(y;\nu)$. 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):[10]

$$\nu\left(F_{t,G}\right) = \nu(F) + t\cdot\int \mathrm{IF}(y;\nu)\cdot d\left(G-F\right)(y) + r\left(t;\nu;G,F\right) \tag{6}$$

where $r\left(t;\nu;G,F\right)$ is a remainder term that converges to zero as t goes to zero at the general rate $o\left(t\right)$. Depending on the functional ν considered, the remainder may converge faster or even be identical to zero. For example, for the mean μ, the variance σ^2 and the quantile q_τ we show in the appendix that $r\left(t;\mu;G,F\right)=0$, $r\left(t;\sigma^2;G,F\right)=o\left(t^2\right)$ and that $r\left(t;q_\tau;G,F\right)=o\left(t\right)$. Also, if $F=G$, then $r\left(t;\nu;F,F\right)=0$ for any t or ν. Thus, the further apart the distributions F and G, the larger should be the remainder term. If we fix ν and t (for example, by making it equal to 1) and allow F and G to be empirical distributions $\widehat{F}$ and $\widehat{G}$, we should expect that the importance of the remainder term to be an empirical question.

Now consider the leading term of equation (6) as an approximation for $\nu\left(G\right)$, that is, for $t=1$:

$$\nu\left(G\right) \approx \nu(F) + \int \mathrm{IF}(y;\nu)\cdot dG\left(y\right). \tag{7}$$

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)

$$\mathrm{RIF}(y;\nu,F) = \nu(F) + \int \mathrm{IF}(y;\nu)\cdot d\Delta_y\left(y\right) = \nu(F) + \mathrm{IF}(y;\nu). \tag{8}$$

Again by a normalization argument, we write $\mathrm{RIF}(y;\nu,F)$ as $\mathrm{RIF}(y;\nu)$. The recentered influence function $\mathrm{RIF}(y;\nu)$ has several interesting properties:

[10] 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 $A'(0)=\int a_1(y)\,d\left(G-F\right)(y)$, where $a_1(y)$ is the influence function, we get the VOM approximation.

Property 1 [***Mean and Variance of the Recentered Influence Function***]***:***
i) the $\mathrm{RIF}(y;\nu)$ *integrates up to the functional of interest* $\nu(F)$

$$\int \mathrm{RIF}(y;\nu)\cdot dF\,(y)=\int (\nu(F)+\mathrm{IF}(y;\nu))\cdot dF\,(y)=\nu(F). \tag{9}$$

ii) the $\mathrm{RIF}(y;\nu)$ *has the same asymptotic variance as the functional* $\nu(F)$

$$\int (\mathrm{RIF}(y;\nu)-\nu(F))^2\cdot dF\,(y)=\int (\mathrm{IF}(y;\nu))^2\cdot dF\,(y)=AV\,(\nu,F) \tag{10}$$

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

Property 2 [***Recentered Influence Function and the Directional Derivative***]***:***
i) the derivative of the functional $\nu\,(F_{t,G})$ *in the direction of the distribution* G *is obtained by integrating up the recentered influence function at* F *over the distributional differences between* G *and* F[11]

$$\frac{\partial\nu\,(F_{t,G})}{\partial t}|_{t=0}=\int \mathrm{RIF}(y;\nu)\cdot d\,(G-F)\,(y)\,. \tag{11}$$

ii) the Von Mises approximation (6) can be written in terms of the $\mathrm{RIF}(y;\nu)$ *as*

$$\nu\,(F_{t,G})=\nu(F)+t\cdot\int \mathrm{RIF}(y;\nu)\cdot d\,(G-F)\,(y)+r\,(t;\nu;G,F) \tag{12}$$

where the remainder term is

$$r\,(t;\nu;G,F)=\int (\mathrm{RIF}(y;\nu,F_{t,G})-\mathrm{RIF}(y;\nu))\cdot dF_{t,G}\,(y)\,.$$

Note that properties 1 *ii*) and 2 *i*) are also shared by the influence function.

3.2 Impact of General Changes in the Distribution of X

We now show that the recentered influence function provides a convenient way of assessing the impact of changes in the covariates on the distributional statistic ν without having to compute the corresponding counterfactual distribution which is, in general, a difficult estimation problem. We first consider general changes in the distribution of covariates,

[11] This property combines (4), (5) and (8) and follows from the fact that densities integrate to one.

from $F_X(x)$ to the counterfactual distribution $G_X(x)$. We then consider the special case of a marginal change from X to $X+t$, and of the policy change $\ell(X)$ introduced in Section 2.

In the presence of covariates X, we can use the law of iterated expectations to express ν in terms of the conditional expectation of $\mathrm{RIF}(y;\nu)$ given X (the RIF-regression function $E[\mathrm{RIF}(Y;\nu)|X=x]$):

Property 3 [***Link between the functional*** $\nu(F)$ ***and the RIF-regression***]***:*** *The RIF-regression* $E[\mathrm{RIF}(Y;\nu)|X=x]$ *integrates up to the functional of interest* $\nu(F)$

$$\nu(F)=\int \mathrm{RIF}(y;\nu)\cdot dF(y)=\int E[\mathrm{RIF}(Y;\nu)|X=x]\cdot dF_X(x) \tag{13}$$

where we have substituted equation (1) into equation (9), and used the fact that $E[\mathrm{RIF}(Y;\nu)|X=x]=\int_y \mathrm{RIF}(y;\nu)\cdot dF_{Y|X}(y|X=x)$. Property 3 is central to our approach. It provides a simple way of writing any functional ν of the distribution as an expectation and, furthermore, to write ν as the mean of the RIF-regression $E[\mathrm{RIF}(Y;\nu)|X]$. Comparing equation (1) and property 3 illustrates how our approach greatly simplifies the modelling of the effect of covariates on distribution statistics. In equation (1), the whole conditional distribution, $F_{Y|X}(y|X=x)$, has to be integrated over the distribution of X to get the unconditional distribution of Y, F.[12] When we are only interested in a specific distribution statistic $\nu(F)$, however, we simply need to integrate over $E[\mathrm{RIF}(Y;\nu)|X]$, which is easily estimated using regression methods.

Looking at property 3 also suggests computing counterfactual values of ν by integrating over a counterfactual distribution of X instead of $F_X(x)$. We now precisely state the main theorem that enables us to use these types of counterfactual manipulations. The theorem shows that the effect (on the functional ν) of a small change in the distribution of covariates from F_X in the direction of G_X is given by integrating up the RIF-regression function with respect to the change in distribution of the covariates, $d(G_X-F_X)$.

[12] This is essentially what Machado and Mata (2005) suggest to do, since they propose estimating the whole conditional distribution by running (conditional) quantile regressions for each and every possible quantile.

Theorem 1 [***Marginal Effect of a Change in the Distribution of X***]*:*

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

The proof, provided in the appendix, starts with equation (11) and builds on a Lemma showing the impact of distributional changes in response to a change in G_X.

Lets' now consider the implications of Theorem 1 for the cases of the policy effect and of the unconditional partial effect introduced in Section 2. Given that $\pi_G(\nu)$ captures the marginal effect of moving the distribution of X from F_X to G_X, it can be used as the leading term of an approximation, just like equation (11) is the leading term of the von Mises approximation (equation (12)). Our first corollary shows how this fact can be used to approximate in the policy effect $\delta_\ell(\nu)$.

Corollary 1 [***Policy Effect***]*: If the policy change from X to $\ell(X)$ can be described as a change in the distribution of covariates, that is, $\ell(X) \sim G_X(x) = F_X(\ell^{-1}(x))$, then, $\delta_\ell(\nu)$, the policy effect on the functional ν consists of the marginal effect of the policy, $\pi_\ell(\nu)$, and a remainder term $r(\nu,G,F)$:*

$$\delta_\ell(\nu) = \nu(G) - \nu(F) = \pi_\ell(\nu) + r(\nu;G,F)$$

where

$$\begin{aligned} \pi_\ell(\nu) &= \int E\left[\mathrm{RIF}(Y;\nu)|X=x\right]\cdot\left(dF_X(\ell^{-1}(x)) - dF_X(x)\right), \ \textit{and} \\ r(\nu;G,F) &= \int \left(E\left[\mathrm{RIF}(Y;\nu,G)|X=x\right] - E\left[\mathrm{RIF}(Y;\nu)|X=x\right]\right)\cdot dG_X(x) \end{aligned}$$

The proof is provided in the appendix. Note that the approximation error $r(\nu;G,F)$ depends on how different the means of $\mathrm{RIF}(Y;\nu)$ and $\mathrm{RIF}(Y;\nu,G)$ are.[13]

The next case is the unconditional partial effect of X on ν, defined as $\alpha(\nu)$ in Section 2. The implicit assumption used is that X is a continuous covariate that is being increased from X to $X+t$. We consider the case where X is discrete in the third corollary below.

[13] We discuss this issue in more detail in Section 6.

Corollary 2 [***Unconditional Partial Effect: Continuous Covariate***]***:*** *Consider increasing a continuous covariate X by t, from X to $X+t$. This change results in the counterfactual distribution $F_t^*(y) = \int F_{Y|X}(y|x) \cdot dF_X(x-t)$. The effect of X on the distributional statistic ν, $\alpha(\nu)$, is*

$$\begin{aligned} \alpha(\nu) &\equiv \lim_{t \downarrow 0} \frac{\nu(F_t^*) - \nu(F)}{t} \\ &= \int \frac{dE[\mathrm{RIF}(Y;\nu)|X=x]}{dx} \cdot dF(x). \end{aligned}$$

The proof is provided in the appendix. The corollary simply states that the effect (on ν) of a small change in covariate X is equal to the average derivative of the recentered influence function with respect to the covariate.[14]

Finally, we consider the case where X is a dummy variable. The manipulation we have in mind here consists of increasing the probability that X is equal to one by a small amount t

Corollary 3 [***Unconditional Partial Effect: Dummy Covariate***]***:*** *Consider the case where X is a discrete (dummy) variable, $X \in \{0,1\}$. Define $P_X \equiv \Pr[X=1]$. Consider an increase from P_X to P_X+t. This results in the counterfactual distribution $F_t^*(y) = F_{Y|X}(y|1) \cdot (P_X+t) + F_{Y|X}(y|0) \cdot (1-P_X-t)$. The effect of a small increase in the probability that $X=1$ is given by*

$$\begin{aligned} \alpha_D(\nu) &\equiv \lim_{t \downarrow 0} \frac{\nu(F_t^*) - \nu(F)}{t} \\ &= E[\mathrm{RIF}(Y;\nu,F)|X=1] - E[\mathrm{RIF}(Y;\nu,F)|X=0] \end{aligned}$$

The proof is, once again, provided in the appendix.

4 Application to Unconditional Quantiles

In this section, we apply the results of Section 3 to the case of quantiles. We first show that, for quantiles, the RIF is a linear transformation of a dichotomous variable indicating whether Y is above or below the quantile. This suggests estimating $E[\mathrm{RIF}(Y;q_\tau)|X=x]$ using probit or logit regressions, or a simple OLS regression (linear probability model). These estimation issues are discussed in detail in the next section. Note that for other

[14] In the case of a multivariate X, the relevant concept is the average partial derivative.

functionals ν besides the quantile, estimation of $E[\mathrm{RIF}(Y;\nu)|X=x]$ may be more appropriately pursued by nonparametric methods. We then look at a number of special cases that help interpret the RIF regressions in terms of the underlying structural model, $Y=h(X,\varepsilon)$, and provide some guidance on the functional form of the RIF regressions. We also show the precise link between the UQPE and the CQPE, which is closely connected to the structural form.

4.1 Recentered Influence Functions for Quantiles

As a benchmark, first consider the case of the mean, $\nu(F)=\mu$. Applying the definition of influence function of (equation (5)) to $\mu=\int y\cdot dF(y)$, we get that $\mathrm{IF}(y;\mu)=y-\mu$, and that $\mathrm{RIF}(y;\mu)=\mu+\mathrm{IF}(y;\mu)=y$. When the VOM linear approximation of equation (12) is applied to the mean, μ, the remainder $r(t;\mu;G,F)$ equals zero since $\mathrm{RIF}(y;\mu)=\mathrm{RIF}(y;\mu,F_{t,G})=y$.

Turning to our application of interest, consider the τ^{th} quantile: $\nu(F)=q_\tau$. Applying the definition of influence function to $q_\tau=\inf_q \Pr[Y\le q]\ge\tau$, it follows that

$$\begin{aligned}\mathrm{IF}(y;q_\tau) &= \frac{\tau-\mathbb{1}\{y\le q_\tau\}}{f_Y(q_\tau)}\\ &= \frac{\mathbb{1}\{y>q_\tau\}}{f_Y(q_\tau)}-\frac{1-\tau}{f_Y(q_\tau)}.\end{aligned}$$

The influence function is simply a dichotomous variable that takes on the value $-(1-\tau)/f_Y(q_\tau)$ when Y is below the quantile q_τ, and $\tau/f_Y(q_\tau)$ when Y is above the quantile q_τ. The recentered influence function is

$$\begin{aligned}\mathrm{RIF}(y;q_\tau) &= q_\tau+\mathrm{IF}(y;q_\tau)\\ &= q_\tau+\frac{\tau-\mathbb{1}\{y\le q_\tau\}}{f_Y(q_\tau)}\\ &= \frac{\mathbb{1}\{y>q_\tau\}}{f_Y(q_\tau)}+q_\tau-\frac{1-\tau}{f_Y(q_\tau)}\\ &= c_{1,\tau}\cdot\mathbb{1}\{y>q_\tau\}+c_{2,\tau}.\end{aligned}$$

where $c_{1,\tau}=1/f_Y(q_\tau)$ and $c_{2,\tau}=q_\tau-c_{1,\tau}\cdot(1-\tau)$. Note that equation (9) implies that the mean of the recentered influence function is the quantile of q_τ itself, and from equation (10) that its variance is $\tau\cdot(1-\tau)/f_Y^2(q_\tau)$.

The main results in Section 3 all involve the conditional expectation of the RIF. In the case of quantiles, we have

$$
\begin{aligned}
E\left[\mathrm{RIF}(Y;q_\tau)|X=x\right] &= c_{1,\tau}\cdot E\left[\mathbb{1}\left\{Y>q_\tau\right\}|X=x\right]+c_{2,\tau} \\
&= c_{1,\tau}\cdot \Pr\left[Y>q_\tau|X=x\right]+c_{2,\tau} \qquad (14)
\end{aligned}
$$

Since the conditional expectation $E\left[\mathrm{RIF}(Y;\nu)|X=x\right]$ is a linear function of $\Pr\left[Y>q_\tau|X=x\right]$, it can be estimated using a standard probability response model (e.g. probit, logit, or linear probability model).[15] The estimated model can then used to compute either the policy effect or the UQPE defined in Corollaries 1 to 3.

Consider the case of the unconditional partial effect (UQPE in the case of quantiles) with continuous regressors in Corollary 2, $\alpha(\tau)$. From Corollary 2 we have

$$
\begin{aligned}
UQPE\left(\tau\right) &= \alpha(\tau)=\int\frac{dE\left[\mathrm{RIF}(Y;q_\tau)|X=x\right]}{dx}\cdot dF_X\left(x\right) \qquad (15)\\
&= \frac{1}{f_Y\left(q_\tau\right)}\cdot\int\frac{d\Pr\left[Y>q_\tau|X=x\right]}{dx}\cdot dF_X\left(x\right) \qquad (16)\\
&= c_{1,\tau}\cdot\int\frac{d\Pr\left[Y>q_\tau|X=x\right]}{dx}\cdot dF_X\left(x\right) \qquad (17)
\end{aligned}
$$

The integral in the above equation is the usual "marginal" effect of the covariates in a probability response model (see, e.g., Wooldridge (2002)).[16] Interestingly, the UQPE for a dummy regressor is also closely linked to a standard marginal effect in a probability response model. In that case, it follows from Corollary 3 that

$$
\begin{aligned}
UQPE\left(\tau\right) &= \alpha_D(\tau)\\
&= \frac{1}{f_Y\left(q_\tau\right)}\cdot\left(\Pr\left[Y>q_\tau|X=1\right]-\Pr\left[Y>q_\tau|X=0\right]\right)\\
&= c_{1,\tau}\cdot\left(\Pr\left[Y>q_\tau|X=1\right]-\Pr\left[Y>q_\tau|X=0\right]\right).
\end{aligned}
$$

At first glance, the fact that the UQPE is closely linked to standard marginal effects in a probability response model is a bit surprising. Consider a particular value y_0 of Y that corresponds to the τ^{th} quantile of the distribution of Y, q_τ. Except for the multiplicative factor $1/f_Y\left(q_\tau\right)$, our results mean that a small increase in X has the same impact on the probability that Y is above y_0, than on the τ^{th} unconditional quantile of Y. In other

[15] The parameters $c_{1,\tau}$ and $c_{2,\tau}$ can be estimated using the sample estimate of q_τ and a kernel density estimate of $f\left(q_\tau\right)$. See Section 5 for more detail.

[16] Note that the marginal effect is often computed as the effect of X on the probability for the "average observation", $\frac{d\Pr[Y\geq q_\tau|X=\overline{x}]}{dx}$. This is how STATA, for example, computes marginal effects. The more appropriate marginal effect here, however, is the average of the marginal effect for each observation.

words, we can transform a probability impact into an unconditional quantile impact by simply multiplying the probability impact by $1/f_Y(q_\tau)$. Roughly speaking, the reason why $1/f_Y(q_\tau)$ provides the right transformation is that the function that transforms probabilities into unconditional quantiles is the inverse of the cumulative distribution function, $F_Y^{-1}(y)$, and the slope of $F_Y^{-1}(y)$ is the inverse of the density, $1/f_Y(q_\tau)$. In essence, the proposed approach enables us to turn a difficult estimation problem (the effect of X on unconditional quantiles of Y) into an easy estimation problem (the effect of X on the probability of being above a certain value of Y).

4.2 The UQPE and the structural form

In Section 2, we first defined the UQPE and the policy effect in terms of the structural form $Y = h(X, \varepsilon)$. We now re-introduce the structural form to show how it is linked to the RIF-regression model, $E[\mathrm{RIF}(Y; q_\tau)|X = x]$. This is useful for interpreting the parameters of the RIF-regression model, and for suggesting possible functional forms for the regression.

We explore these issues using three specific examples of the structural form, and then discuss the link between the UQPE and the structural form in the most general case where $h(\cdot)$ is completely unrestricted (except for the assumption of monotonicity in ε). Even in that general case, we show that the UQPE can be written as a weighted average of the CQPE, which is closely connected to the structural form, for different quantiles and values of X.

4.2.1 Case 1: Linear, additively separable model

We start with the simplest linear model $Y = h(X, \varepsilon) = X^\intercal\beta + \varepsilon$. As discussed in Section 2 2, we limit ourselves to the case where X and ε are independent. The linear form of the model implies that a small change t in a covariate X_j simply shifts the location of the distribution of Y by $\beta_j \cdot t$, but leaves all other features of the distribution unchanged. As a result, the UQPE for any quantile is equal to β_j. While β could be estimated using a standard OLS regression in this simple case, it is nonetheless useful to see how it could also be estimated using our proposed approach.

For the sake of simplicity, assume that ε follows a distribution F_ε. Note that this parametric assumption will turn out to be irrelevant in this very restrictive case of a linear structural function.

The resulting probability response model is[17]

$$\Pr\left[Y > q_\tau | X = x\right] = \Pr\left[\varepsilon > q_\tau - x^\intercal \beta\right] = 1 - F_\varepsilon\left(q_\tau - x^\intercal \beta\right).$$

For example, when ε is normally distributed, the probability response model is a standard probit model. Taking derivatives with respect to X_j yield

$$\frac{d \Pr\left[Y > q_\tau | X = x\right]}{d X_j} = \beta_j \cdot f_\varepsilon\left(q_\tau - x^\intercal \beta\right),$$

where f_ε is the density of ε, and the marginal effects are obtained by integrating over the distribution of X

$$\int \frac{d \Pr\left[Y > q_\tau | X = x\right]}{d x_j} \cdot d F_X\left(x\right) = \beta_j \cdot E\left[f_\varepsilon\left(q_\tau - X^\intercal \beta\right)\right],$$

where the expectation on the right hand side is taken over the distribution of X. As it turns out, the expression inside the expectation operator is simply the conditional density of Y evaluated at $Y = q_\tau$:

$$f_{Y|X}(q_\tau | X = x) = f_\varepsilon\left(q_\tau - x^\intercal \beta\right).$$

It follows that

$$\int \frac{d \Pr\left[Y > q_\tau | X = x\right]}{d x_j} \cdot d F_X\left(x\right) = \beta_j \cdot E[f_{Y|X}(q_\tau | X = x)] = \beta_j \cdot f_Y(q_\tau),$$

and that (by substituting back in equation (16)) the UQPE is indeed equal to β_j

$$UQPE_j\left(\tau\right) = \frac{1}{f\left(q_\tau\right)} \cdot \beta_j \cdot f_Y(q_\tau) = \beta_j.$$

4.2.2 Case 2: Non-linear, additively separable model

A simple extension of the linear model is the index model $h(X, \varepsilon) = h\left(X^\intercal \beta + \varepsilon\right)$, where h is differentiable and monotonic. When h is non-linear, a small change t in a covariate X_j does not simply shift the location of the distribution of Y, and the UQPE is no longer equal to β. One nice feature of the model, however, is that it yields that same probability

[17] Since q_τ is just a constant, it can be absorbed in the usual constant term.

response model as in Case 1. We have

$$\begin{aligned}\Pr[Y > q_\tau | X = x] &= \Pr\left[\varepsilon > h^{-1}(q_\tau) - x^\intercal\beta\right] \\ &= 1 - F_\varepsilon\left(h^{-1}(q_\tau) - x^\intercal\beta\right),\end{aligned}$$

The average marginal effects are now

$$\int \frac{d\Pr[Y > q_\tau | X = x]}{dx_j} \cdot dF_X(x) = \beta_j \cdot E\left[f_\varepsilon\left(h^{-1}(q_\tau) - X^\intercal\beta\right)\right],$$

and the UQPE is

$$UQPE_j(\tau) = \beta_j \cdot \frac{E\left[f_\varepsilon\left(h^{-1}(q_\tau) - X^\intercal\beta\right)\right]}{f_Y(q_\tau)} = \beta_j \cdot h'\left(h^{-1}(q_\tau)\right),$$

where the last equality follows from the fact that

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

Since $h'\left(h^{-1}(q_\tau)\right)$ depends on q_τ, it follows that the UQPE is proportional, but not equal, to the underlying structural parameter β. Also, the UQPE does not depend on the distribution of ε. The intuition for this result is simple. From Case 1, we know that the effect of X_j on the τ^{th} quantile of the index $X^\intercal\beta + \varepsilon$ is β_j. But since Y and $X^\intercal\beta + \varepsilon$ are linked by a rank preserving transformation $h(\cdot)$, the effect on the τ^{th} quantile of Y corresponds to the effect on the τ^{th} quantile of the index $X^\intercal\beta + \varepsilon$ times the slope of the transformation function evaluated at this point, $h'\left(h^{-1}(q_\tau)\right)$.[18]

4.2.3 Case 3: Linear, separable, but heteroskedastic model

A more standard model used in economics is the linear, but heteroskedastic model $h(X, \varepsilon) = X^\intercal\beta + \sigma(X) \cdot \varepsilon$, where X and ε are still independent, but where $Var(Y|X) = \sigma^2(X)$.[19] The special case where $\sigma(X) = X^\intercal\psi$ has the interesting implication that the

[18] Note that the UQPE could also be obtained by estimating the index model using a flexible form for $h(\cdot)$ (see, for example, Fortin and Lemieux (1998)). Estimating a flexible form for $h(\cdot)$ and taking derivatives is more difficult, however, than just computing average marginal effects and dividing them by the density $f(q_\tau)$. More importantly, since the index model is very restrictive, it is important to use a more general approach that is robust to the specifics of the structural form $h(X, \varepsilon)$.

[19] There is no loss in generality in assuming that $Var(\varepsilon) = 1$.

conventional conditional quantile regression functions are also linear in X, an assumption typically used in practice. To see this, consider the τ^{th} conditional quantile of Y, $Q_\tau[Y|X=x]$

$$Q_\tau[Y|X=x] = Q_\tau[X^\intercal\beta + (X^\intercal\psi)\cdot\varepsilon|X=x] = x^\intercal(\beta + Q_\tau[\varepsilon]\cdot\psi),$$

where $Q_\tau[\varepsilon]$ is the τ^{th} quantile of ε.[20] This particular specification of $h(X,\varepsilon)$ can also be related to the quantile structural function (QSF) of Imbens and Newey (2005). In the case where ε is a scalar (as here), Imbens and Newey define the QSF as $Q_\tau[Y|X=x] = h(x, Q_\tau[\varepsilon])$, which simply corresponds to $x^\intercal(\beta + Q_\tau[\varepsilon]\cdot\psi)$ is the special case considered here.

The implied probability response model is the heteroskedastic model

$$\Pr[Y > q_\tau|X=x] = \Pr\left[\varepsilon > \frac{-(x^\intercal\beta - q_\tau)}{x^\intercal\psi}\right] = 1 - F_\varepsilon\left(\frac{q_\tau - x^\intercal\beta}{x^\intercal\psi}\right). \tag{18}$$

As is well known (e.g. Wooldridge, 2002), introducing heteroskedasticity greatly complicates the interpretation of the structural parameters (β and ψ here). The problem is that even if β_j and ψ_j are both positive, a change in X increases both the numerator and the denominator in equation (18), which has an ambiguous effect on the probability. In other words, it is no longer possible to express the marginal effects as simple functions of the conditional mean parameter, β, as we did in Cases 1 and 2.

Strictly speaking, after imposing a parametric assumption on the dsitribution of ε, such as $\varepsilon \sim N(0,1)$, one could take this particular model at face value and estimate the implied non-linear probit model using maximum likelihood, and then compute the probit marginal effects to get the UQPE. A more practical solution, however, is to estimate a more standard flexible probability response model and compute the marginal effects. We propose such a non-parametric approach in Section 5.

4.2.4 General case

One potential drawback of estimating a flexible probability response model, however, is that we then lose the tight connection between the UQPE and the underlying structural

[20] For example, if ε is normal, the median $Q_{.5}[\varepsilon]$ is zero and the conditional median regression is $Q_{.5}[Y|X=x] = x^\intercal\beta$. Similarly, the 90^{th} quantile $Q_{.9}[\varepsilon]$ is 1.28 and the corresponding regression for the 90^{th} quantile is $Q_{.9}[Y|X=x] = x^\intercal(\beta + 1.28\cdot\psi)$. Note also that this specific model yields a highly restricted set of quantile regressions in a multivariate setting, since the vector of parameters ψ is mutiplied by a single factor $Q_\tau[\varepsilon]$. Allowing for a more general specification would only make the results below even more cumbersome.

parameters highlighted, for example, in Case 2 above. Fortunately, it is still possible to draw a useful connection between the UQPE and the underlying structural form, even in a very general case.

By analogy with the UQPE, consider the conditional quantile partial effect (CQPE), which represents the effect of a small change of X on the conditional quantile of Y

$$\begin{aligned} CQPE\left(\tau,x\right) &\equiv \lim_{t\downarrow 0}\frac{Q_{\tau}\left[h(X+t,\varepsilon)|X=x\right]-Q_{\tau}\left[Y|X=x\right]}{t} \\ &= \frac{\partial Q_{\tau}\left[h(X,\varepsilon)|X=x\right]}{\partial x} \end{aligned}$$

The CQPE is the derivative of the conditional quantile regression with respect to X. In the standard case of a linear quantile regression, the CQPE simply corresponds to the quantile regression coefficient. Using the definition of the QSF, we can also express the CQPE as

$$CQPE\left(\tau,x\right)=\frac{\partial Q_{\tau}\left[h(X,\varepsilon)|X=x\right]}{\partial x}=\frac{\partial h(x,Q_{\tau}\left[\varepsilon\right])}{\partial x}.$$

Before we establish the link between the UQPE and the CQPE, let us define the following three auxiliary functions. The first one, $\omega_{\tau}:\mathcal{X}\rightarrow\mathbb{R}^{+}$, will be used as a weighting function and is basically the ratio between the conditional density given $X=x$, and the unconditional density:

$$\omega_{\tau}\left(x\right)\equiv\frac{f_{Y|X}\left(q_{\tau}|x\right)}{f_{Y}\left(q_{\tau}\right)}.$$

The second function, $\varepsilon_{\tau}:\mathcal{X}\rightarrow\mathbb{R}$, is the inverse h function $h^{-1}\left(\cdot,q_{\tau}\right)$, which exists under the assumption that h is monotonic in ε:

$$\varepsilon_{\tau}\left(x\right)\equiv h^{-1}\left(x,q_{\tau}\right).$$

Finally, the third function, $s_{\tau}:\mathcal{X}\rightarrow(0,1)$, can be thought as a "matching" function that shows where the unconditinal quantile q_{τ} falls in the conditional distribution of Y:

$$s_{\tau}\left(x\right)\equiv\left\{\top:Q_{\top}\left[Y|X=x\right]=q_{\tau}\right\}=F_{Y|X}\left(q_{\tau}|X=x\right).$$

We can now state our general result on the link between the $UQPE$ and the $CQPE$

Proposition 1 [***UQPE and its relation to the structural form***]:
i) Assuming that the structural form $Y=h(X,\varepsilon)$ is monotonic in ε and that X and ε

are independent, the parameter $UQPE\left(\tau\right)$ will be:

$$UQPE\left(\tau\right) = E\left[\omega_{\tau}\left(X\right) \cdot \frac{\partial h\left(X, \varepsilon_{\tau}(X)\right)}{\partial x}\right]$$

ii) We can also represent $UQPE\left(\tau\right)$ as a weighted average of $CQPE\left(s_{\tau}(x), x\right)$:

$$UQPE\left(\tau\right) = E\left[\omega_{\tau}\left(X\right) \cdot CQPE\left(s_{\tau}(X), X\right)\right].$$

The proof is provided in the Appendix. Under the hypothesis that X and ε are independent and h monotonic in ε, we may invoke the results by Matzkin (2003) that guarantee that both the distribution of ε and the link function h will be non-parametrically identified. Thus, we know that under the independence assumption, the parameters $UQPE\left(\tau\right)$ are identified.

The proposition shows formally that conditional quantiles do not average up to their respective unconditional quantile, i.e. $UQPE\left(\tau\right) \neq E\left[CQPE\left(\tau, X\right)\right]$. For example, the mean conditional median is not equal to the unconditional median. Instead, the proposition shows that $UQPE\left(\tau\right)$ is equal to a weighted average (over the distribution of X) of the $CQPE$ at the $s_{\tau}(X)$ conditional quantile corresponding to the τ^{th} unconditional quantile of the distribution of Y, q_{τ}. This is better illustrated with a simple example. Suppose that we are looking at the $UQPE$ for the median, $UQPE(.5)$. If X has a positive effect on Y, then the overall median $q_{.5}$ may correspond, for example, to the 30^{th} quantile for observations with a high value of X, but to the 70^{th} quantile for observations with low values of X. In terms of the $s_{\tau}(\cdot)$ function, we have $s_{.5}(X = high) = .3$ and $s_{.5}(X = low) = .7$. Thus, $UQPE(.5)$ is an average of the $CQPE$ at the 70^{th} and 30^{th} quantiles, respectively, which may arbitrarily differ from the $CQPE$ at the median.

More generally, whether or not the $UQPE\left(\tau\right)$ is "close" to $CQPE\left(\tau, X\right)$ depends on the functional form of $h(X, \varepsilon)$ and on the distribution of X (and $\omega_{\tau}\left(X\right)$). In Case 1 above, the CQPE is the same for all quantiles ($CQPE\left(\tau, X\right) = \beta$ for all τ). Since UQPE is a weighted average of the CQPE's, it trivially follows that $UQPE\left(\tau\right) = CQPE\left(\tau, X\right) = \beta$. Another special case is when the function $s_{\tau}(X)$ does not vary very much and is more or less equal to τ for all values of X. This would tend to happen when the model has little explanatory power, i.e. when most of the variation in Y is in the "residuals". In the simple example above, this may mean, for instance, that $s_{.5}(X = high) = .49$ and $s_{.5}(X = low) = .51$. By a simple continuity argument, $CQPE\left(.49, X\right)$ and $CQPE\left(.51, X\right)$ are very close to each other (and to $CQPE\,(.5,X)$) and have

$$UQPE\,(\tau)=E\left[\omega_{\tau}\,(X)\cdot CQPE\,(s_{\tau}(X),X)\right]\approx E\left[\omega_{\tau}\,(X)\cdot CQPE\,(\tau,X)\right]. \tag{19}$$

When quantile regressions are linear ($Q_{\tau}\,[Y|X=x]=x^{\intercal}\beta_{\tau}$), we get that $CQPE\,(\tau,X)=CQPE\,(\tau)=\beta_{\tau}$, and the right hand side of equation (19) is equal to $CQPE\,(\tau)$. It follows that $UQPE\,(\tau)\approx CQPE\,(\tau)$. These issues will be explored further in the context of the empirical examples in Section 6.

5 Estimation

In this section, we discuss the estimation of the $UQPE\,(\tau)$ and of the policy effect $\delta_{\ell}\,(\tau)$ (as approximated by the parameter $\pi_{\ell}\,(\tau)$ in Corollary 1) using unconditional quantile regressions. Before discussing the regression estimators, however, we first consider the estimation of the recentered influence function, which depends on some unknown objects (the quantile and the density) of the marginal distribution of Y. We thus start by presenting formally the estimators for q_{τ}, $f_{Y}\,(\cdot)$, and $\mathrm{RIF}(y;q_{\tau})$.

As discussed in Section 4, estimating the RIF-regression for quantiles, called $m_{\tau}\,(x)$ in this section, is closely linked to the estimation of a probability response model since

$$m_{\tau}\,(x)\equiv E\left[\mathrm{RIF}(Y;q_{\tau})|X=x\right]=c_{1,\tau}\cdot\Pr\left[Y>q_{\tau}|X=x\right]+c_{2,\tau}.$$

It follows from equation (15) that:

$$UQPE\,(\tau)\equiv\int\frac{dm_{\tau}\,(x)}{dx}\cdot dF_{X}\,(x)=c_{1,\tau}\cdot\int\frac{d\Pr\left[Y>q_{\tau}|X=x\right]}{dx}\cdot dF_{X}\,(x)\,.$$

For the sake of convenience, we also define the random variable T_{τ}, where

$$T_{\tau}=1\!\!1\left\{Y>q_{\tau}\right\}.$$

The remainder of the section present estimators of $UQPE\,(\tau)$ and $\pi_{\ell}\,(\tau)$ based on three specific regression methods: (i) RIF-OLS, (ii) RIF-Logit, and (iii) RIF-nonparametric, where the latter estimation method is based on a nonparametric version of the logit model. Since the $UQPE\,(\tau)$ is a function of average marginal effects (Section 4), all three estimators may yield relative accurate estimates of $UQPE\,(\tau)$ given that marginal effects from a linear probability model (RIF-OLS) or a logit (RIF-Logit or RIF-nonparametric) are often very similar, in practice. This issue will be explored in more detail in the empirical section (Section 6). The main asymptotic results for the estimators of $UQPE(\tau)$ for each one of these estimation methods can be found in the appendix.[21]

Though we discuss the estimation of RIF-regressions for the case of quantiles, the estimation approach can be easily extended to a general functional of the unconditional distribution of Y, $\nu(F_Y)$. In that general case, the parameters of interest $\alpha(\nu)$ and $\delta_\ell(\nu)$ would involve estimation of both $E[\text{RIF}(Y;\nu)|X=x]$ and the expectation of its derivative, $E[dE[\text{RIF}(Y;\nu)|X]/dX]$ (in the case of the unconditional partial effect, $\alpha(\nu)$). As in the case of quantiles, one could, in principle, estimate those two objects using either an OLS regression or nonparametric methods (e.g. a series estimator). Estimators suggested in the literature on average derivative estimation (e.g., Hardle and Stoker, 1989) could be used to estimate $E[dE[\text{RIF}(Y;\nu)|X]/dX]$.

5.1 Recentered Influence Function and its components

In order to estimate $UQPE(\tau)$ and $\pi_\ell(\tau)$ we first have to obtain the estimated recentered influence functions. Since T_τ is a non-observable random variable that depends on the true unconditional quantile q_τ, we use a feasible version of that variable

$$\widehat{T}_\tau = \mathbb{1}\{Y > \widehat{q}_\tau\}.$$

The corresponding feasible version of the RIF is

$$\begin{aligned}\widehat{\text{RIF}}(Y;\widehat{q}_\tau) &= \widehat{q}_\tau + \frac{\tau - \mathbb{1}\{Y \leq \widehat{q}_\tau\}}{\widehat{f}_Y(\widehat{q}_\tau)} \\ &= \widehat{c}_{1,\tau} \cdot \widehat{T}_\tau + \widehat{c}_{2,\tau},\end{aligned}$$

which also involves two unknown quantities to be estimated, $\widehat{q}_\tau$ and $\widehat{f}_Y(\widehat{q}_\tau)$. The estimator of the τ^{th} population quantile of the marginal distribution of Y is $\widehat{q}_\tau$, the usual τ^{th} sample quantile, which can be represented, using Koenker and Bassett (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 em-

[21] Estimators of the parameter $\pi_G(\tau)$ will depend on the particular choice of policy change and therefore will not have their asymptotic properties analyzed here.

pirical section we propose using the Gaussian kernel with associated optimal bandwidth. The actual requirements for the kernel and the bandwidth are described in the asymptotics section of the appendix. 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:

$$\widehat{f}_Y(y) = \frac{1}{N \cdot b_Y} \cdot \sum_{i=1}^{N} K_Y\left(\frac{Y_i - y}{b_Y}\right). \tag{20}$$

Finally, the parameters $c_{1,\tau}$ and $c_{2,\tau}$ are estimated as $\widehat{c}_{1,\tau} = 1/\widehat{f}_Y(\widehat{q}_\tau)$ and $\widehat{c}_{2,\tau} = \widehat{q}_\tau - \widehat{c}_{1,\tau} \cdot (1-\tau)$, respectively.

5.2 Three estimation methods

5.2.1 RIF-**OLS Regression**

The first estimator for $UQPE(\tau)$ and $\pi_\ell(\tau)$ using a simple linear regression. As in the familiar OLS regression, we implicitly 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 covariates. If the linearity assumption seems inappropriate in particular applications, one can always turn to a more flexible estimation method proposed next. Moreover, OLS is known to produce the linear function of covariates that minimizes the specification error.

The RIF-OLS estimator for $m_\tau(x)$ is

$$\widehat{m}_{\tau,\mathrm{RIF}-OLS}(x) = x^\intercal \cdot \widehat{\gamma}_\tau,$$

where $\widehat{\gamma}_\tau$ is also the estimator for the derivative $dm_\tau(x)/dx$. The estimated coefficient vector is simply a projection coefficient

$$\widehat{\gamma}_\tau = \left(\sum_{i=1}^{N} X_i \cdot X_i^\intercal\right)^{-1} \cdot \sum_{i=1}^{N} X_i \cdot \widehat{\mathrm{RIF}}(Y_i;\widehat{q}_\tau). \tag{21}$$

As mentioned eralier, the RIF-OLS estimator is closely connected to a linear probability model for $\mathbb{1}\{Y \le \widehat{q}_\tau\}$. The projection coefficients $\widehat{\gamma}_\tau$ (except for the constant) are equal to the coefficients in a linear probability model divided by the rescaling factor $\widehat{f}_Y(\widehat{q}_\tau)$.

The estimators for $UQPE(\tau)$ and $\pi_\ell(\tau)$ are

$$\widehat{UQPE}_{\mathrm{RIF}-OLS}(\tau) = \widehat{\gamma}_\tau$$

$$\widehat{\pi}_{\ell,\mathrm{RIF}-OLS} = \widehat{\gamma}_\tau^\intercal \cdot \frac{1}{N} \cdot \sum_{i=1}^{N} \left(\ell(X_i) - X_i\right)$$

5.2.2 RIF-Logit Regression

The second estimator exploits the fact that the regression model is closely connected to a probability response model since $m_\tau(x) = c_{1,\tau} \cdot \Pr[T_\tau = 1|X = x] + c_{2,\tau}$. Assuming a logistic model

$$\Pr[T_\tau = 1|X = x] = \Lambda(x^\intercal \theta_\tau),$$

where is $\Lambda(\cdot)$ is the logistic CDF, we can estimate θ_τ by maximum likelihood by replacing T_τ by its feasible counterpart $\widehat{T}_\tau$:

$$\widehat{\theta}_\tau = \arg\max_{\theta_\tau} \sum_{i}^{N} \widehat{T}_{\tau,i} \cdot X_i^\intercal \theta_\tau + \log\left(1 - \Lambda\left(X_i^\intercal \theta_\tau\right)\right)$$

The main advantage of the logit model over the linear specification for $m_\tau(x)$ is that it allows heterogenous marginal effects, that is, for $dm_\tau(x)/dx$ to depend on x:

$$\frac{dm_\tau(x)}{dx} = c_{1,\tau} \cdot \frac{d\Pr[T_\tau = 1|X = x]}{dx} = c_{1,\tau} \cdot \theta_\tau \cdot \Lambda(x^\intercal \theta_\tau) \cdot (1 - \Lambda(x^\intercal \theta_\tau)).$$

Thus, we propose estimating $UQPE(\tau)$ and $\pi_\ell(\tau)$ as

$$\begin{aligned}
\widehat{UQPE}_{\mathrm{RIF}-logit}(\tau) &= \widehat{c}_{1,\tau} \cdot \widehat{\theta}_\tau \cdot \frac{1}{N} \cdot \sum_{i=1}^{N} \Lambda\left(X_i^\intercal \widehat{\theta}_\tau\right) \cdot \left(1 - \Lambda\left(X_i^\intercal \widehat{\theta}_\tau\right)\right) \\
\widehat{\pi}_{\ell,\mathrm{RIF}-logit}(\tau) &= \widehat{c}_{1,\tau} \cdot \frac{1}{N} \cdot \sum_{i=1}^{N} \left(\Lambda\left(\ell(X_i)^\intercal \widehat{\theta}_\tau\right) - \Lambda\left(X_i^\intercal \widehat{\theta}_\tau\right)\right)
\end{aligned}$$

5.2.3 Nonparametric-RIF Regression (RIF-NP)

The third estimator does not make any functional form assumption about $\Pr[Y > q_\tau|X = x]$. We use the method proposed by Hirano, Imbens and Ridder (2003) to estimate a probability response model nonparametrically by means of a polynomial approximation of the log-odds ratio of $\Pr[Y > q_\tau | X = x]$.[22] The specifics of the problem are the following. We will estimate a vector $\rho_K(\tau)$ of length $K(\tau)$ by finding the solution to the problem

$$\widehat{\rho}_K(\tau) = \arg\max_{\rho_K(\tau)} \sum_i^N \widehat{T}_{\tau,i} \cdot H_{K(\tau)}(X_i)^\intercal \rho_K(\tau) + \log\left(1 - \Lambda\left(H_{K(\tau)}(X_i)^\intercal \rho_K(\tau)\right)\right)$$

where $H_{K(\tau)}(x) = [H_{K(\tau),j}(x)] \quad (j = 1, ..., K(\tau))$, a vector of length $K(\tau)$ of polynomial functions of $x \in \mathcal{X}$ satisfying the following properties: (i) $H_{K(\tau)}: \mathcal{X} \to \mathbb{R}^{K(\tau)}$; (ii) $H_{K(\tau),1}(x) = 1$, and (iii) if $K(\tau) > (n+1)^r$, then $H_{K(\tau)}(x)$ includes all polynomials up order n.[23] In what follows, we assume that $K(\tau)$ is a function of the sample size N such that $K(\tau) \to \infty$ as $N \to \infty$.[24] Our estimate of $\Pr[T_\tau = 1 | X = x]$ is now

$$\widehat{p}_{K,\tau}(x) \equiv \Lambda\left(H_{K(\tau)}(x)^\intercal \widehat{\rho}_K(\tau)\right).$$

Thus, we propose estimating $UQPE(\tau)$ and $\pi_\ell(\tau)$ as

$$\begin{aligned}
\widehat{UQPE}_{\text{RIF}-NP}(\tau) &= \widehat{c}_{1,\tau} \cdot \widehat{\rho}_K(\tau) \cdot \frac{1}{N} \cdot \sum_{i=1}^N \frac{dH_{K(\tau)}(X_i)^\intercal}{dx} \cdot \widehat{p}_{K,\tau}(X_i) \cdot (1 - \widehat{p}_{K,\tau}(X_i)) \\
\widehat{\pi}_{\ell,\text{RIF}-NP}(\tau) &= \widehat{c}_{1,\tau} \cdot \frac{1}{N} \cdot \sum_{i=1}^N (\widehat{p}_{K,\tau}(\ell(X_i)) - \widehat{p}_{K,\tau}(X_i))
\end{aligned}$$

It is interesting to see how this nonparametric approach relates to the previous method based on a logit specification. If $H_{K(\tau)}(x) = x$ for all x, that is, if $H_{K(\tau)}(\cdot)$ is the identity function, then the two methods coincide. Thus, our nonparametric approach generalizes the RIF-logit method. The nonparametric approach can also be interpreted as a flexible logit model that incorporates not only a linear function inside the logistic, but a richer polynomial that includes interactions, squares and cubics.

6 Empirical Applications

In this section, we present two empirical applications to illustrate how the unconditional quantile regressions work in practice, and how the results compare to standard (condi-

[22] The log odds ratio of $\Pr[Y > q_\tau | X = x]$ is equal to $\log(\Pr[Y > q_\tau | X = x] / (1 - \Pr[Y > q_\tau | X = x]))$.

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

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

tional) quantile regressions. In the first application, we revisit the analysis of birthweight of Koenker and Hallock (2001), where they use quantile regressions to 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. The second application looks at effect on unions on male wages, which is well known to be different at different points of the wage distribution (see, for example, Chamberlain, 1994, and Card, 1996).

6.1 Determinants of Birthweight

In the case of infant birthweight, "just looking at conditional means" is too restrictive from a public health perspective, since we may be particularly concerned with the lower tail of the birthweight distribution, and in particular with cases that fall below the "low birthweight" threshold of 2500 grams. In this setting, the limitation of quantile regressions as a distributional tool is that the low birthweight threshold may well fall at different quantiles depending on the characteristics of the mother. For example, the 10th quantile for black high school dropout mothers who also smoke (at 2183 grams) is well below the low birthweight threshold of 2500 grams. By contrast, the 10th quantile for white college educated mothers who do not smoke (at 2880 grams) is well above the low birthweight threshold. 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.

Proposition 1 shows that the precise link between the effect of covariates on conditional (CQPE) and unconditional (UQPE) quantiles depends on a complicated mix of factors. In practice, the CQPE and the UQPE turn out to be fairly similar in the case of birthweight. For instance, Table 1 compares standard OLS estimates to the RIF-OLS regressions and the conventional (conditional) quantile regressions at the 10^{th}, 50^{th}, and 90^{th} percentiles of the birthweight distribution. While estimates tend to vary substantially across the different quantiles, the difference between RIF-OLS and quantile regression coefficients tends to be small, relative to standard errors.

Figure 1a also shows that the point estimates from conditional and unconditional (both RIF-OLS and RIF-logit) quantile regressions are generally very close and rarely statistically different for the various covariates considered.[25] This reflects the fact that, despite a large sample of 198,377 observations, the standard errors are quite large, a

[25] Confidence intervals are not reported in figure 1, but they almost always overlap for conditional and unconditional quantile regression estimates.

pattern that can also be found in Figure 4 of Koenker and Hallock (2001).[26] In other words, the covariates do not seem to be explaining much of the overall variation in birthweight. This is confirmed in Figure A1, which shows that covariates (gender in this example) explain little of the variation in birthweight since the conditional and unconditional distributions are very similar (both look like Gaussian distributions slightly shifted one from another). This corresponds to the case discussed after Proposition 1 where the function $s_\tau(X)$ does not vary very much, and is more or less equal to τ for all values of X. As a result, it is not very surprising that the UQPE and CQPE are quite close to each other.

6.2 Unions and Wage Inequality

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. These ambiguous effects are compounded by the fact that the union wage gap generally declines as a function of the (observed) skill level.[27]

Table 2 reports the RIF-OLS estimates for the 10^{th}, 50^{th} and 90^{th} quantiles using a large sample of U.S. males from the 1983-85 Outgoing Rotation group (ORG) supplement of the Current Population Survey.[28] 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 result that unions have different effects at different points of the wage distribution. Note that the effects are very precisely estimated 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

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

[27] See Card, Lemieux, and Riddell (2004) for a detailed discussion and survey of the distributional effects of unions.

[28] 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). We have run the models for different time periods but only present the 1983-85 results here.

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, which is very different from the unconditional quantile regressions estimates. The difference between conditional and unconditional quantile regression estimates is illustrated in detail in Figure 2, which plots both conditional and unconditional quantile regression estimates for each covariate at 19 different quantiles (from the 5^{th} to the 95^{th}). Both the RIF-OLS and RIF-logit ($\widehat{UQPE}_{\text{RIF}-logit}(\tau)$) estimates are reported. While the estimated union effect is very different for conditional and unconditional quantiles, results obtained using RIF-OLS or RIF-logit regressions are very similar. This confirms the "common wisdom" in empirical work that marginal effects from a linear probability model (RIF-OLS) or a logit (RIF-OLS) are very similar.

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. 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 unlike the birthweight example, we now have a case where there are large and important differences between the $UQPE$ and the $CQPE$. This is consistent with the fact that the conditional and unconditional distribution of log wages are more dissimilar than in the case of birthweight. Figure A2 shows that 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. By contrast, the distribution of wages for non-union workers is closer to a normal distribution, though it also has a mass point in the lower tail at the minimum wage.

Figure 3 illustrates more formally that the RIF-OLS regressions appear to be providing very robust estimates of the underlying parameter of interest, the $UQPE$. The first two panels of Figure 3 compares confidence intervals of RIF-OLS estimates to those obtained by estimating conditional quantile regressions (first panel) or by computing the marginal effects from a logit regression (RIF-logit).[29] These two figures show that un-

[29] 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.

conditional regression estimates are robust to the estimation method used in the sense that confidence intervals are hardly distinguishable from each other. This conclusion is reinforced by the third panel of Figure 3, which shows that using the fully nonparametric estimator (RIF-NP) yields estimates that are virtually identical to those obtained with the RIF-logit or RIF-OLS estimator.[30] This is in sharp contrast with the very big difference in confidence intervals obtained when comparing RIF-OLS estimates with conditional quantile regression estimates.

The last panel of Figure 3 shows, however, that even if the density is precisely estimated, the choice of the bandwidth does matter for some of the estimates 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.

Having established that the RIF-OLS method works very well in practice, we return to the important question how well it can approximate the effect of "larger changes" in covariates such as those contemplated for the policy effect of Corollary 1. To assess the importance of the approximation error, $r(\nu; G, F)$, we conduct a small experiment looking at the effect of unions 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 in 1983-85).

Since we only have a dummy covariate, it is also straightforward to compute an "exact" effect of unionization by simply changing the proportion of union workers, and recomputing the various quantiles in this "reweighted" sample.[31] The resulting estimates are the diamonds reported in Figure 4a to 4g. Generally speaking, the RIF-OLS estimates are remarkably close to the "exact" estimates, even for large changes in union-

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

[31] 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.

ization (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 4 (Figure 4h) 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 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:

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

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 $\Delta \widetilde{Var(w)}$ while the exact effect is $\Delta Var(w)$ from above. The approximation error is, thus, the second order term $\Delta Var(w) - \Delta \widetilde{Var(w)} = -(\Delta U)^2 \cdot D^2$. It corresponds to the difference between the straight line and the diamonds in Figure 4g. 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 regression method to estimate the effect of explanatory variables on the unconditional quantiles of an outcome variable. The proposed unconditional quantile regression method consists of running a regression of the (recentered) influence function of the unconditional quantile of the outcome 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. We show how standard partial effects, that we call unconditional quantile partial effects (UQPE) for the problem studied here, can be estimated using our regression approach. The regression estimates can also be used to approximate the effect of a more general change in the distribution of covariates (policy effect) on unconditional quantiles of the outcome variable. We propose three different regression estimators based on a standard OLS regression (RIF-OLS, where the recentered influence function is the dependent variable), a logit regression (RIF-logit), and a nonparametric logit regression (RIF-OLS).

We show in the empirical section that our estimators are very easy to use in practice, and that RIF-OLS, RIF-logit, and RIF-NP all yield very similar estimates of the UQPE in the applications considered. We present two applications that illustrate well the differences between conditional and unconditional quantile regressions. In the first application, the analysis of infant birthweight of Koenker and Hallock (2001), conditional and unconditional quantile regression estimates are very close to each other. In the second application, the effect of unions on the wage distribution, the results are more strikingly different. While traditional quantile regressions indicate that unions have a positive effect on wages even at the top quantiles of the wage distribution, we actually find a strong negative effect of unions at the highest quantile of the wage distribution. We also show that our unconditional quantile regressions approximate very well the effect of larger changes in the rate of unionization on unconditional quantiles of the wage distribution.

Another important advantage of the proposed method is that it can be easily generalized to other distributional statistics such as the Gini or the Theil coefficient. All that is required to do so is to compute the recentered influence function for these statistics, and run regression of the resulting RIF on the covariates. We discuss in a companion paper (Firpo, Fortin, and Lemieux, 2005) how our regression method can be used to generalize traditional Oaxaca-Blinder decompositions (for means) to any distributional statistic.

One limitation of the proposed regression method is the assumption that the covariates, X, are independent of unobservables, ε, in the general model $Y = h(X, \varepsilon)$ for the outcome variable, Y. While the independence assumption combined with a flexible form of $h(X, \varepsilon)$ still allows for rich distributional effects of X on Y (such as heteroskedasticity in a standard regression model), it is nonetheless highly restrictive in many problems of economic interest, such as the effect of schooling on the distribution of wages. As is well known, there are good economic reasons why schooling may be correlated with unobservables such as ability (Card, 2001). We plan to show in future work how the independence assumption can be relaxed when instrumental variables are available for the endogenous covariates, and how consistent estimates of the UQPE can be obtained by adding a control function in the unconditional quantile regressions.

8 Appendix

Proof of Theorem 1: Let us begin with

LEMMA 1 *Since, by definition, G_Y only changes in response to a change in G_X, the counterfactual mixing distribution which represents the distribution of y resulting from a change t away from F_X in the direction of the alternative probability distribution G_X is equal to the original mixing distribution $F_{Y,t,G_Y} = (1-t) \cdot F_Y + t \cdot G_Y$.*

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

with $F_{X,t,G_X} = (1-t) \cdot F_X + t \cdot G_X$.

Proof: We have

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

Thus

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

■

Now 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:

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

Now, note that by lemma 1

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

thus,

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

■

Proof of Corollary 1:

Consider the leading term of equation (6) as an approximation for $\nu\left(G_Y\right)$, that is, for $t=1$:

$$\nu\left(G_Y\right)\approx\nu(F_Y)+\int \mathrm{IF}(y;\nu,F_Y)\cdot dG_Y(y)\,.$$

Because by property i) the $\mathrm{RIF}(y;\nu,F)$ integrates up to the functional $\nu(F)$, we have

$$\begin{aligned}\nu\left(F_{t,G_Y}\right) &= \int \mathrm{RIF}(y;\nu,F_{t,G_Y})\cdot dF_{t,G_Y}\left(y\right)\\ \nu\left(G_Y\right) &= \int \mathrm{RIF}(y;\nu,G_Y)\cdot dG_Y\left(y\right)\end{aligned}$$

and

$$\begin{aligned}\nu\left(F_{t,G_Y}\right) &= \nu(F_Y)+t\cdot\int \mathrm{IF}(y;\nu,F_Y)\cdot d\left(G_Y-F_Y\right)(y)+r\left(t;\nu;G_Y,F_Y\right)\\ &= \nu(F_Y)+t\cdot\int \mathrm{RIF}(y;\nu,F_Y)\cdot d\left(G_Y-F_Y\right)(y)+r\left(t;\nu;G_Y,F_Y\right)\\ &= \nu(F_Y)+t\cdot\left(\nu\left(G_Y\right)-\nu\left(F_Y\right)\right)\\ &\quad +t\cdot\int\left(\mathrm{RIF}(y;\nu,F_Y)-\mathrm{RIF}(y;\nu,G_Y)\right)\cdot dG_Y\left(y\right)+r\left(t;\nu;G_Y,F_Y\right)\end{aligned}$$

Rearranging terms and taking the limit $t\downarrow 0$

$$\begin{aligned}\lim_{t\downarrow 0}\frac{\nu\left(F_{t,G_Y}\right)-\nu(F_Y)}{t} &= \nu\left(G_Y\right)-\nu\left(F_Y\right)\\ &\quad +\int\left(\mathrm{RIF}(y;\nu,F_Y)-\mathrm{RIF}(y;\nu,G_Y)\right)\cdot dG\left(y\right)+\frac{r\left(t;\nu;G_Y,F_Y\right)}{t}\\ &= \int \mathrm{RIF}(y;\nu,F)\cdot d\left(G-F\right)(y)\\ &= \nu\left(G_Y\right)-\nu\left(F_Y\right)+\int\left(\mathrm{RIF}(y;\nu,F)-\mathrm{RIF}(y;\nu,G_Y)\right)\cdot dG_Y\left(y\right)\end{aligned}$$

then by Theorem 1 and Lemma 1

$$\begin{aligned}&\nu\left(G_Y\right)-\nu\left(F_Y\right)=\pi(\nu,G_X)+\int\left(\mathrm{RIF}(y;\nu,G_Y)-\mathrm{RIF}(y;\nu,F_Y)\right)\cdot dG_Y\left(y\right)\\ &\quad =\pi(\nu,\ell(X))+\int\left(E\left[\mathrm{RIF}(Y;\nu,G_Y)|X=x\right]-E\left[\mathrm{RIF}(Y;\nu,F_Y)|X=x\right]\right)\cdot dg_X(x)\end{aligned}$$

■

Proof of Corollary 2:

Consider an increase t in the variable X. This yields the variable $Z=X+t$, where the density of Z is $f_X(z-t)$. Consider the resulting counterfactual distribution $F^*_{Y,t}$ of Y:

$$\begin{aligned}
F^*_{Y,t}(y) &= \int F_{Y|X}(y|x)\cdot f_X(x-t)\cdot dx \\
&= \int F_{Y|X}(y|x)\cdot f_X(x)\cdot dx - t\cdot\left(\int F_{Y|X}(y|x)\cdot\frac{f'_X(x)}{f_X(x)}\cdot f_X(x)\cdot dx + o(|t|)\right) \\
&= F_Y(y) + t\cdot\left(\int F_{Y|X}(y|x)\cdot l_X(x)\cdot f_X(x)\cdot dx + o(|t|)\right)
\end{aligned}$$

where the second line is obtained using a first order expansion, and where

$$l_X(x) = -\frac{d\ln(f(x))}{dx} = -\frac{f'_X(x)}{f_X(x)}.$$

Define

$$g_X(x) = f_X(x) + l_X(x)\cdot f_X(x).$$

Since

$$G_Y(y) = \int F_{Y|X}(y|x)\cdot g_X(x)\cdot dx,$$

it follows that

$$G_Y(y) = F_Y(y) + \int F_{Y|X}(y|x)\cdot l_X(x)\cdot f_X(x)\cdot dx.$$

Since we can write

$$F^*_{Y,t}(y) = F_Y(y) + t\cdot((G_Y(y) - F_Y(y)) + o(|t|)) = F_{Y,t\cdot G_Y}(y) + t\cdot o(|t|),$$

if follows that

$$\begin{aligned}
\alpha(\nu) &\equiv \lim_{t\downarrow 0}\frac{\nu\left(F^*_{Y,t}\right) - \nu(F_Y)}{t} \\
&= \lim_{t\downarrow 0}\left(\frac{\nu\left(F_{Y,t\cdot G_Y} + t\cdot o(|t|)\right) - \nu(F_Y)}{t}\right) \\
&= \lim_{t\downarrow 0}\left(\frac{\nu\left(F_{Y,t\cdot G_Y}\right) - \nu(F_Y)}{t}\right)
\end{aligned}$$

Using Theorem 1, it follows that

$$\begin{aligned}\beta(\nu) &= \pi(\nu, G_X) = \int E\left[\mathrm{RIF}(Y;\nu,F)|X=x\right]\cdot d(G_X - F_X)(x) \\ &= \int E\left[\mathrm{RIF}(Y;\nu,F)|X=x\right]\cdot l_X(x)\cdot f_X(x)\cdot dx\end{aligned}$$

By partial integration

$$\begin{aligned}&\int E\left[\mathrm{RIF}(Y;\nu,F)|X=x\right]\cdot l_X(x)\cdot f_X(x)\cdot dx \\ =& -\int E\left[\mathrm{RIF}(Y;\nu,F)|X=x\right]\cdot f_X'(x)\cdot dx \\ =& \int \frac{dE\left[\mathrm{RIF}(Y;\nu,F)|X=x\right]}{dx}\cdot f_X(x)\cdot dx \\ =& \int \frac{dE\left[\mathrm{RIF}(Y;\nu,F)|X=x\right]}{dx}\cdot f_X(x)\cdot dx\end{aligned}$$

hence

$$\alpha(\nu) = \int \frac{dE\left[\mathrm{RIF}(Y;\nu,F)|X=x\right]}{dx}\cdot f_X(x)\cdot dx$$

■

Proof of Corollary 3:
Consider an increase t in the probability P_X that $X=1$. The resulting distribution of X, $F^*_{X,t}$, is given by $dF^*_{X,t}(X=0) = dF_X(X=0) - t$ and $dF^*_{X,t}(X=1) = dF_X(X=1)+t$, where $dF_X(X=0) = (1-P_X)$ and $dF_X(X=1) = P_X$. Since X is a dummy variable, we have

$$F_Y(y) = \int F_{Y|X}(y|x)\cdot dF_X(x) = (1-P_X)\cdot F_{Y|X}(y|X=0) + P_X\cdot F_{Y|X}(y|X=1)$$

while the counterfactual distribution $F^*_{Y,t}$ of Y is

$$\begin{aligned}F^*_{Y,t}(y) &= (1-P_X-t)\cdot F_{Y|X}(y|x=0) + (P_X+t)\cdot F_{Y|X}(y|x=1) \\ &= F_Y(y) + t\cdot\left[F_{Y|X}(y|X=1) - F_{Y|X}(y|X=0)\right] \\ &= F_Y(y) + t\cdot\left[G_Y(y) - F_Y(y)\right] \\ &= F_{Y,t\cdot G_Y}(y)\end{aligned}$$

where

$$
\begin{aligned}
G_Y(y) &\equiv F_{Y|X}(y|X=1) - F_{Y|X}(y|X=0) + F_Y(y) \\
&= F_{Y|X}(y|X=1) - F_{Y|X}(y|X=0) + (1-P_X)\cdot F_{Y|X}(y|X=0) + P_X \cdot F_{Y|X}(y|X=1) \\
&= (1+P_X)\cdot F_{Y|X}(y|X=1) - P_X \cdot F_{Y|X}(y|X=0) \\
&= \int F_{Y|X}(y|x)\cdot dG_X(x)
\end{aligned}
$$

and where $dG_X(X=0) = -P_X$ and $dG_X(X=1) = 1+P_X$. The definition of the marginal effect is

$$
\begin{aligned}
\alpha_D(\nu) &\equiv \lim_{t\downarrow 0} \frac{\nu\left(F^*_{Y,t}\right) - \nu(F_Y)}{t} \\
&= \lim_{t\downarrow 0} \frac{\nu\left(F_{Y,t\cdot G_Y}\right) - \nu(F_Y)}{t}
\end{aligned}
$$

Using Theorem 1, it follows that

$$
\begin{aligned}
\alpha_D(\nu) &= \pi(\nu, G_X) = \int E\left[\mathrm{RIF}(Y;\nu,F)|X=x\right]\cdot d(G_X - F_X)(x) \\
&= E\left[\mathrm{RIF}(Y;\nu,F)|X=0\right]\cdot d(G_X - F_X)(x=0) + E\left[\mathrm{RIF}(Y;\nu,F)|X=1\right]\cdot d(G_X - F_X)(x=1) \\
&= E\left[\mathrm{RIF}(Y;\nu,F)|X=0\right]\cdot(-P_X - (1-P_X)) + E\left[\mathrm{RIF}(Y;\nu,F)|X=1\right]\cdot((1-P_X) - P_X) \\
&= E\left[\mathrm{RIF}(Y;\nu,F)|X=1\right] - E\left[\mathrm{RIF}(Y;\nu,F)|X=0\right]
\end{aligned}
$$

, where $P_X = dF_X(X=1)$. Consider an increase in the probability that $X=1$ from P_X to P_X+t. This results in the counterfactual distribution $F^*_{Y,t}(y) = \int F_{Y|X}(y|x)\cdot dF^*_{X,t}(x)$, where $dF^*_{X,t}(0) = dF_X(0) - t$ and $dF^*_{X,t}(1) = dF_X(1) + t$.

■

Derivation of the Influence Function of the Mean:

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

$\blacksquare$

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:

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

Thus:

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

$\blacksquare$

Proof of Proposition 1:

$i)$ Starting from equation (16)

$$UQPE(\tau) = -\frac{1}{f_Y(q_\tau)} \cdot \int \frac{d\Pr[Y \leq q_\tau | X = x]}{dx} \cdot dF_X(x),$$

and assuming that the structural form $Y = h(X, \varepsilon)$ is monotonic in ε, so that $\varepsilon_\tau(x) = h^{-1}(x, q_\tau)$, we can write

$$\Pr[Y \leq q_\tau | X = x] = \Pr[\varepsilon \leq \varepsilon_\tau(x) | X = x] = F_{\varepsilon|X}(\varepsilon_\tau; x).$$

Taking the derivative with respect to x, we get

$$d\frac{F_{Y|X}(q_\tau | X = x)}{dx} = f_{\varepsilon|X}(\varepsilon_\tau; x) \cdot \frac{\partial h^{-1}(x, q_\tau)}{\partial x} + \frac{\partial F_{\varepsilon|X}(\cdot; x)}{\partial x}.$$

Defining

$$H(x, \varepsilon_\tau, q_\tau) = h(x, \varepsilon_\tau) - q_\tau$$

then:

$$\frac{\partial h^{-1}(x,q_\tau)}{\partial x} = -\frac{\frac{\partial H(x,\varepsilon_\tau,q_\tau)}{\partial x}}{\frac{\partial H(x,\varepsilon_\tau,q_\tau)}{\partial \varepsilon_\tau}} = -\frac{\frac{\partial h(x,\varepsilon_\tau)}{\partial x}}{\frac{\partial h(x,\varepsilon_\tau)}{\partial \varepsilon_\tau}} = -\frac{\partial h(x,\varepsilon_\tau)}{\partial x} \cdot \left(\frac{\partial h(x,\varepsilon)}{\partial \varepsilon}|_{\varepsilon=\varepsilon_\tau}\right)^{-1}$$

$$\frac{\partial h^{-1}(x,q_\tau)}{\partial q_\tau} = -\frac{\frac{\partial H(x,\varepsilon_\tau,q_\tau)}{\partial q_\tau}}{\frac{\partial H(x,\varepsilon_\tau,q_\tau)}{\partial \varepsilon_\tau}} = \left(\frac{\partial h(x,\varepsilon)}{\partial \varepsilon}|_{\varepsilon=\varepsilon_\tau}\right)^{-1}$$

so that:

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

Substituting in

$$UQPE(\tau) = -\left(f_Y(q_\tau)\right)^{-1} \cdot \int d\frac{F_{Y|X}(q_\tau|X=x)}{dx} \cdot dF_X(x)$$

$$= -\left(f_Y(q_\tau)\right)^{-1} \cdot \int \left(f_{\varepsilon|X}(\varepsilon_\tau; x) \cdot \frac{\partial h(x,\varepsilon_\tau)}{\partial x} \cdot \left(\frac{\partial h(x,\varepsilon)}{\partial \varepsilon}|_{\varepsilon=\varepsilon_\tau}\right)^{-1} - \frac{\partial F_{\varepsilon|X}(\cdot; x)}{\partial x}\right) \cdot dF_X(x)$$

$$= -\left(f_Y(q_\tau)\right)^{-1} \cdot E\left[\frac{\partial h(X,\varepsilon_\tau)}{\partial x} \cdot f_{Y|X}(q_\tau|X) - \frac{\partial F_{\varepsilon|X}(\cdot; X)}{\partial x}\right]$$

Given that ε and X are assumed independent, then $\partial F_{\varepsilon|X}(\cdot|x)/\partial x_j = 0$ and $f_{\varepsilon|X}(\cdot) = f_\varepsilon(\cdot)$, the expression simplifies to

$$UQPE(\tau) = E\left[\frac{\partial h(X,\varepsilon_\tau(X))}{\partial x} \cdot \frac{f_{Y|X}(q_\tau|X)}{f_Y(q_\tau)}\right]$$

ii) Let define the parameter of the conditional quantile regression

$$CQPE(\tau^c, x) = d\frac{Q_{Y|X}(x; \tau^c)}{dx}$$

where τ^c denote the quantile of the conditional distribution: $\tau^c = F_{Y|X}\left(Q_{Y|X}(x;\tau^c)\,|X=x\right)$. Since $Y = h(X,\varepsilon)$ is monotonic is ε

$$h^{-1}\left(x, Q_{Y|X}(x;\tau^c)\right) = \varepsilon_{\tau^c}(x) \quad \text{or} \quad Q_{Y|X}(x;\tau^c) = h\left(x, \varepsilon_{\tau^c}(x)\right)$$

thus

$$d\frac{Q_{Y|X}(x;\tau^c)}{dx} = \frac{\partial h(x,\varepsilon_{\tau^c})}{\partial x} + \frac{\partial h(x,\varepsilon)}{\partial \varepsilon}|_{\varepsilon=\varepsilon_{\tau^c}} \cdot d\frac{\varepsilon_{\tau^c}(x)}{dx}$$

but if independence between X and ε is assumed, the last term vanishes and

$$CQPE(\tau^c, x) = d\frac{Q_{Y|X}(x;\tau^c)}{dx} = \frac{\partial h(x,\varepsilon_{\tau^c})}{\partial x}$$

Letting $s(x,\tau) = \{\tau^c : Q_{Y|X}(x;\tau^c) = q_\tau\}$, we can write:

$$UQPE(\tau) = E\left[CQPE(s(X,\tau),X) \cdot \frac{f_{Y|X}(q_\tau|X)}{f_Y(q_\tau)}\right]$$

References

David H. Autor, L.F. Katz, 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.

Card, D. (2001): Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems, *Econometrica*, 69, 1127–60.

Card, D., T. Lemieux, and W.C. Riddell (2004): Unions and Wage Inequality, *Journal of Labor Research*, 25, 519–62.

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

CHESHER, A. (2003): Identification in Nonseparable Models, *Econometrica*, 71, 1401–1444.

DEATON, A., AND S. NG (1998): Parametric and Nonparametric Approaches to Price and Tax Reform, *Journal of 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 and Labor Relations Review*, 50, 629–651.

FLORENS, J.P., J. J. HECKMAN, C. MEGHIR, AND E. VYTLACIL (2003): Instrumental Variables, Local Instrumental Variables and Control Functions, IDEI Working Paper.

FORTIN, N.M., S. FIRPO, AND T. LEMIEUX (2005): Decomposing Wage Distributions: Estimation and Inference, Unpublished Manuscript, UBC and PUC-Rio.

FORTIN, N.M., AND T. LEMIEUX (1998): Rank Regressions, Wage Distributions, and the Gender Gap, *Journal of Human Resources*, 33, 610–643.

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

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

Imbens, G.W., and W.K. Newey (2005): Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity, Unpublished Manuscript, MIT and Berkeley.

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, 461–98.

Machado, A.F., and J. Mata. (2005): Counterfactual decomposition of changes in wage distributions using quantile regression, *Journal of Applied Econometrics*, 20, 445–465.

Matzkin, R.L. (2003): Nonparametric Estimation of Nonadditive Random Functions, *Econometrica*, 71, 1339–1375.

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.

Stock, J.H. (1989): Nonparametric Policy Analysis, *Journal of the American Statistical Association*, 89, 567–575.

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

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

Table 1: Comparing OLS, Conditional Quantile Regressions (CQR) and Unconditional Quantile Regressions (UQR), Birthweight Model (Koenker and Hallock, 2001)

	OLS	10th centile		50th centile		90th centile	
		UQR	CQR	UQR	CQR	UQR	CQR
Boy	108.867	64.126	67.749	120.147	118.872	143.710	134.691
	(2.418)	(4.827)	(4.625)	(2.688)	(3.613)	(3.932)	(3.613)
Married	60.426	85.505	75.129	52.064	51.584	48.923	52.510
	(3.250)	(7.113)	(6.345)	(3.624)	(3.461)	(4.850)	(4.841)
Black	-198.931	-281.568	-238.387	-173.713	-174.072	-133.556	-164.536
	(3.577)	(8.623)	(6.852)	(3.922)	(3.810)	(4.828)	(5.365)
Mother's Age	36.392	50.536	44.713	34.857	32.150	29.132	32.643
	(1.996)	(4.300)	(3.848)	(2.205)	(2.125)	(3.196)	(3.002)
Mother's Age^2	-0.547	-0.888	-0.762	-0.505	-0.461	-0.350	-0.416
	(0.035)	(0.074)	(0.067)	(0.038)	(0.037)	(0.057)	(0.052)
<u>Mother's Education</u>							
High school	15.140	24.038	13.805	15.767	13.293	5.226	15.855
	(3.757)	(8.374)	(7.191)	(4.135)	(4.001)	(5.454)	(5.617)
Some college	31.210	50.088	33.061	33.268	30.690	15.396	24.522
	(4.187)	(8.918)	(8.034)	(4.645)	(4.459)	(6.426)	(6.292)
College	36.648	61.596	53.715	42.736	35.970	17.153	14.290
	(4.508)	(9.216)	(8.695)	(5.015)	(4.801)	(7.284)	(6.785)
No Prenatal	-183.742	-386.957	-309.842	-111.221	-145.072	-10.767	-83.174
	(13.585)	(39.048)	(25.887)	(13.439)	(14.464)	(17.191)	(20.274)
Prenatal Second	12.047	20.675	18.339	0.836	-0.062	12.847	7.202
	(3.772)	(8.109)	(7.239)	(4.167)	(4.017)	(5.778)	(5.657)
Prenatal Third	30.605	60.692	68.382	-5.581	-0.659	-13.023	-9.816
	(8.305)	(18.131)	(15.928)	(9.119)	(8.844)	(11.433)	(12.394)
Smoker	-167.933	-205.904	-164.017	-166.276	-167.468	-136.094	-169.148
	(6.288)	(15.516)	(11.819)	(6.756)	(6.696)	(7.737)	(9.687)
Cigarettes	-3.695	-6.742	-4.796	-3.478	-3.581	-2.747	-3.169
per day	(0.447)	(1.155)	(0.829)	(0.476)	(0.476)	(0.513)	(0.701)
Mother's Weight	10.158	19.044	21.184	5.934	5.978	0.801	1.125
Gain	(0.303)	(0.666)	(0.568)	(0.345)	(0.323)	(0.530)	(0.447)
Mother's Weight	-0.019	-0.123	-0.146	0.020	0.024	0.102	0.008
Gain^2	(0.000)	(0.008)	(0.008)	(0.005)	(0.005)	(0.008)	(0.006)
Constant	2447.741	1545.573	1583.787	2574.531	2600.281	3305.305	3216.766
	(27.808)	(60.725)	(53.727)	(30.620)	(29.611)	(43.567)	(41.686)

Table 2: Comparing OLS, Conditional Quantile Regressions (CQR) and Unconditional Quantile Regressions (UQR), 1983-85 CPS data 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)
<u>Education</u>							
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	-0.19	-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)
<u>Experience</u>							
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	-0.36	-0.279	-0.377	-0.319
	(0.004)	(0.005)	(0.005)	(0.006)	(0.004)	(0.008)	(0.006)
10-14	-0.149	-0.04	-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	-0.06	-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 1. Unconditional and Conditional Quantile Regressions Estimates for the Birthweight Model

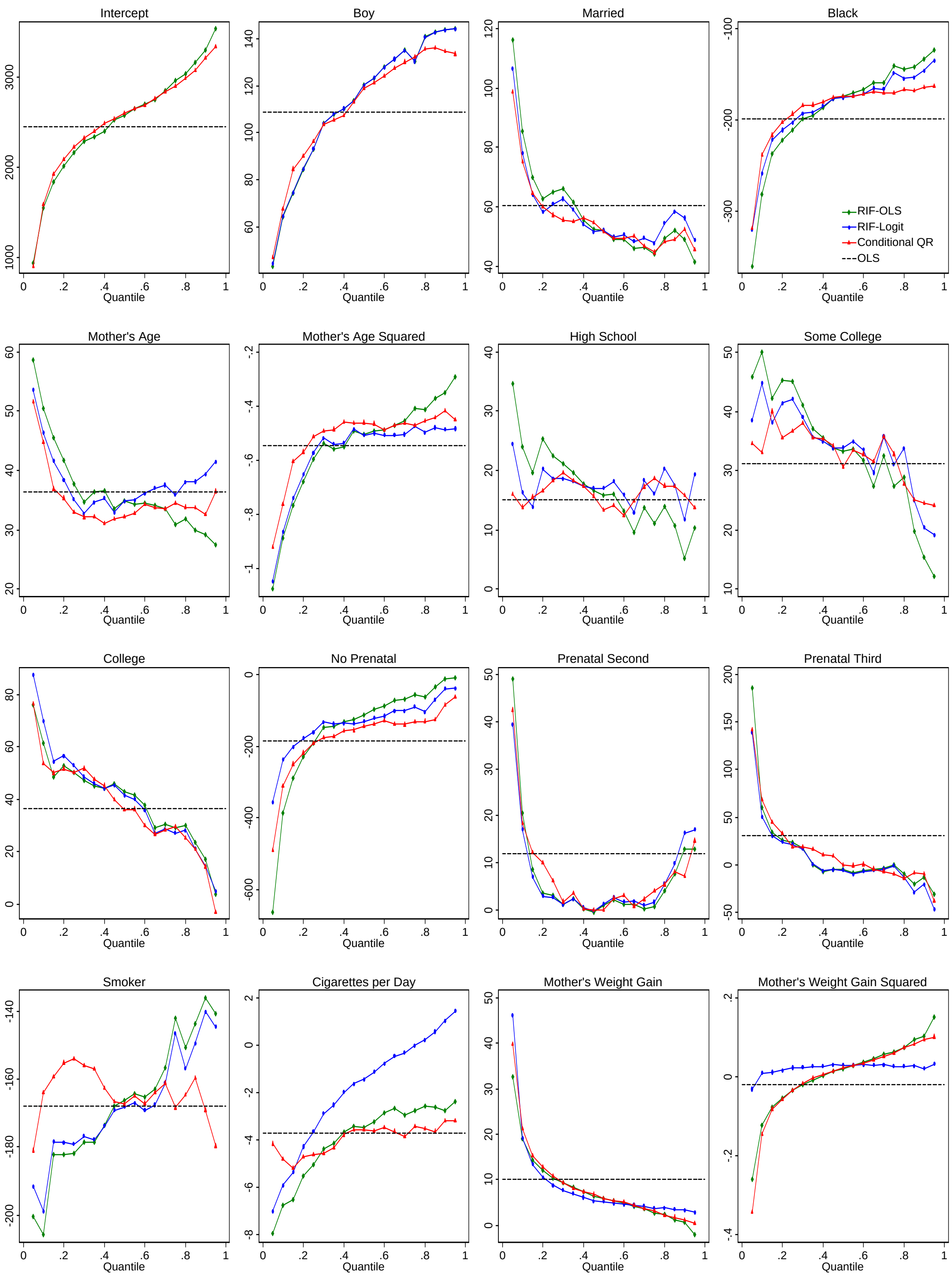

Figure 2. Unconditional and Conditional Quantile Estimates for the Log Wages Model, Men 1983-1985

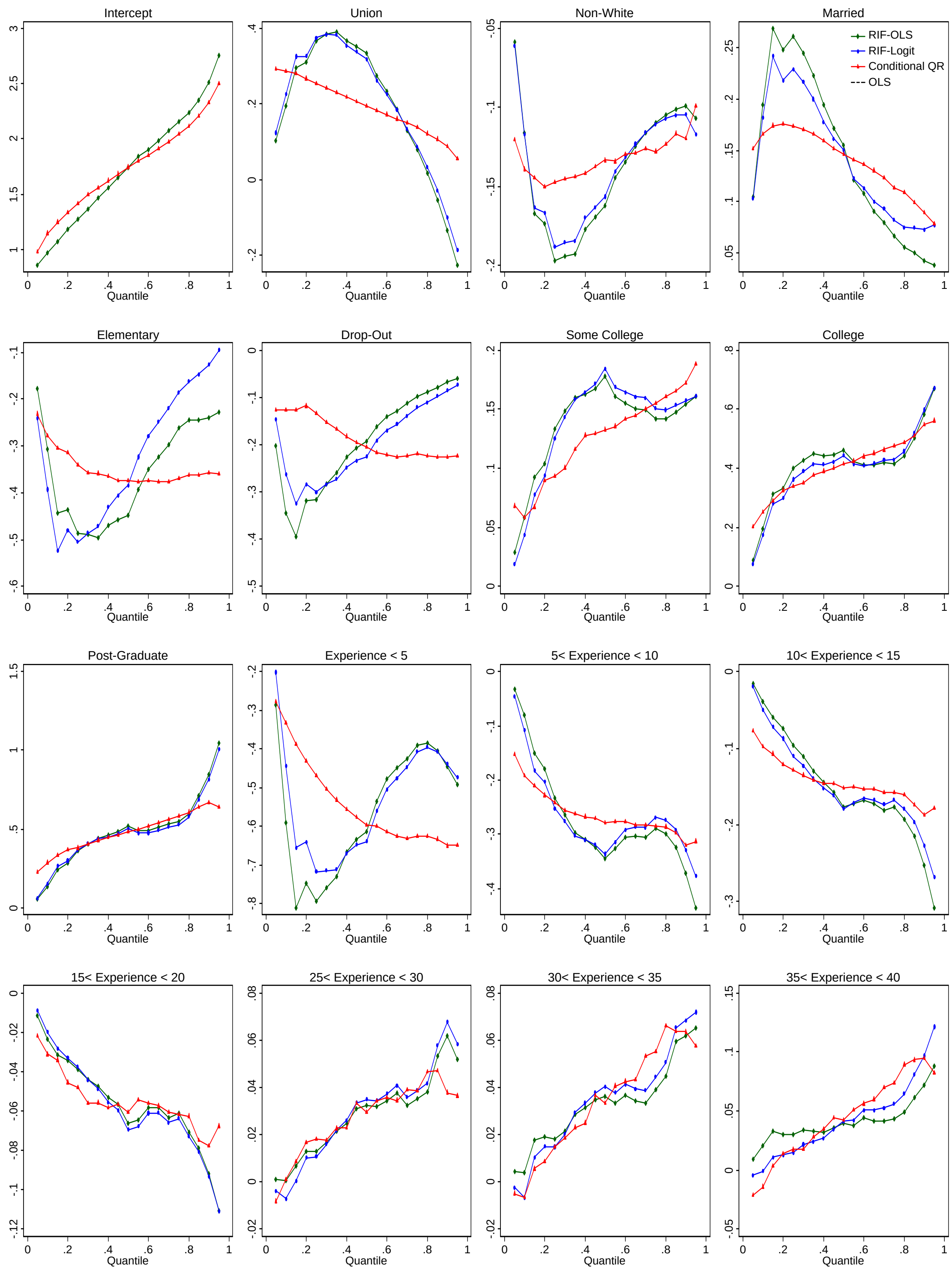

Figure 3. Sensitivity of Unconditional Estimates for the Log Wages Model, Men 1983-1985

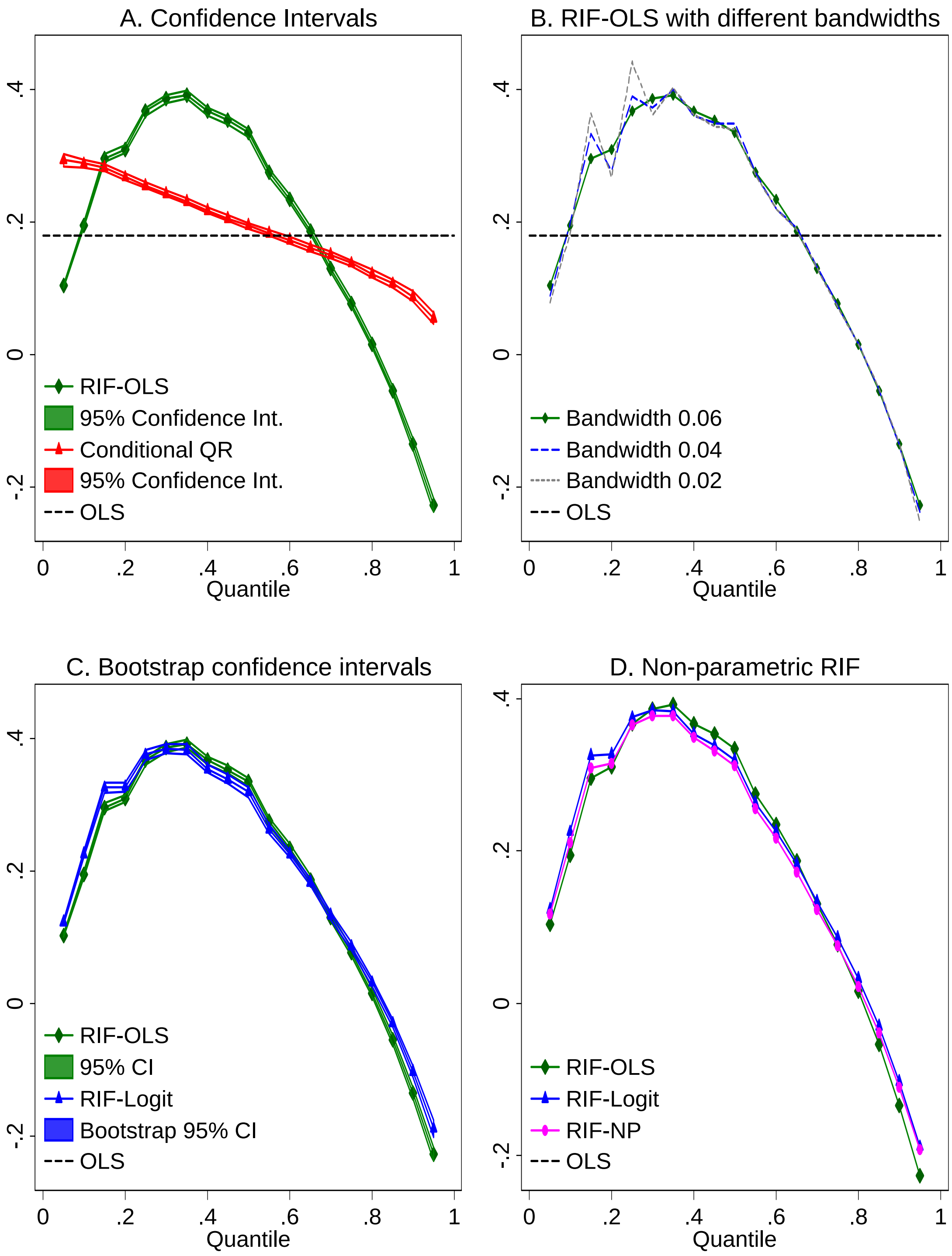

Figure 4: Approximation error (relative to reweighting) when predicting the effect of changes in the unionization rate using the unconditional quantile regression

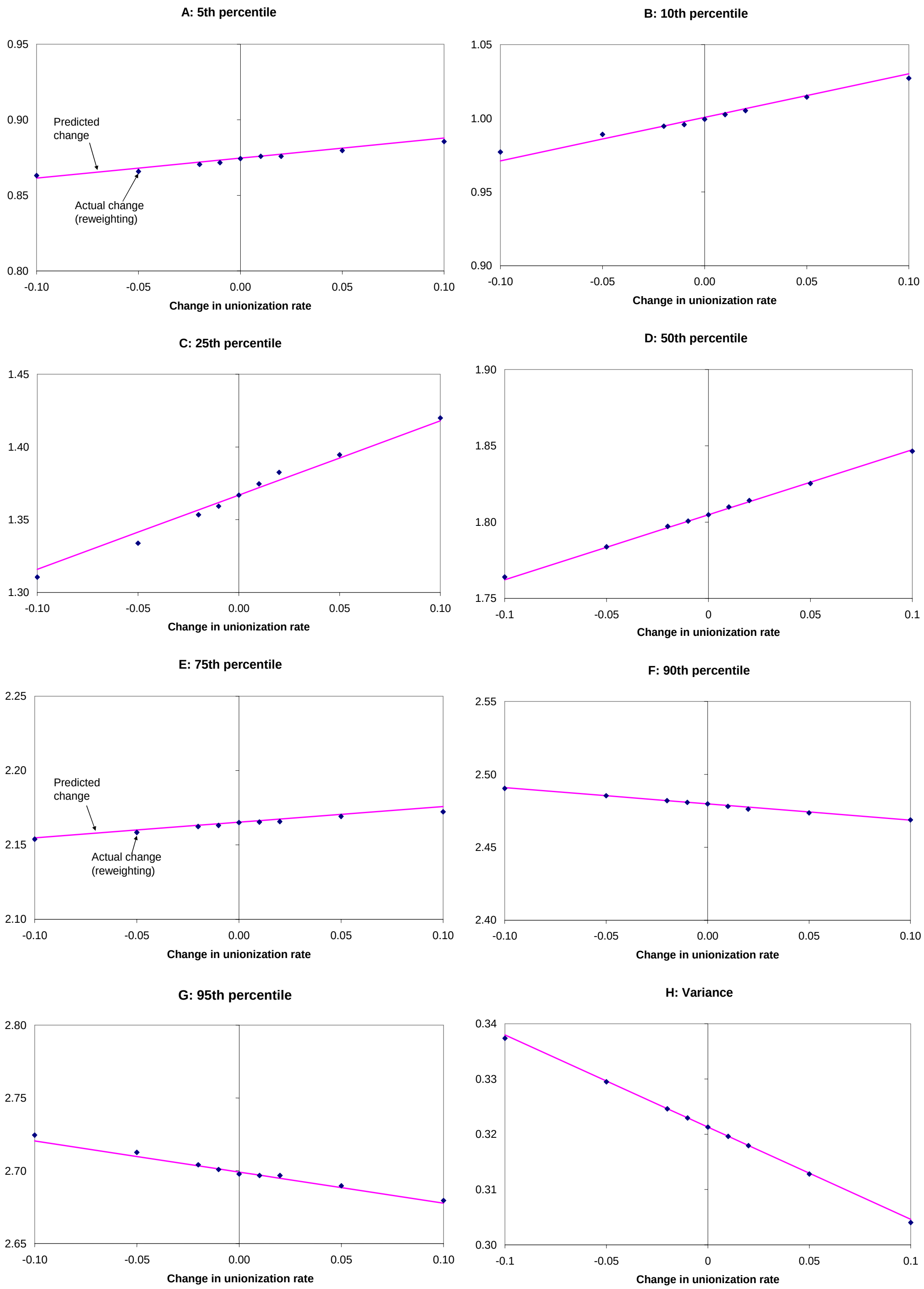

Figure A1. Probability Density Functions of Birthweight

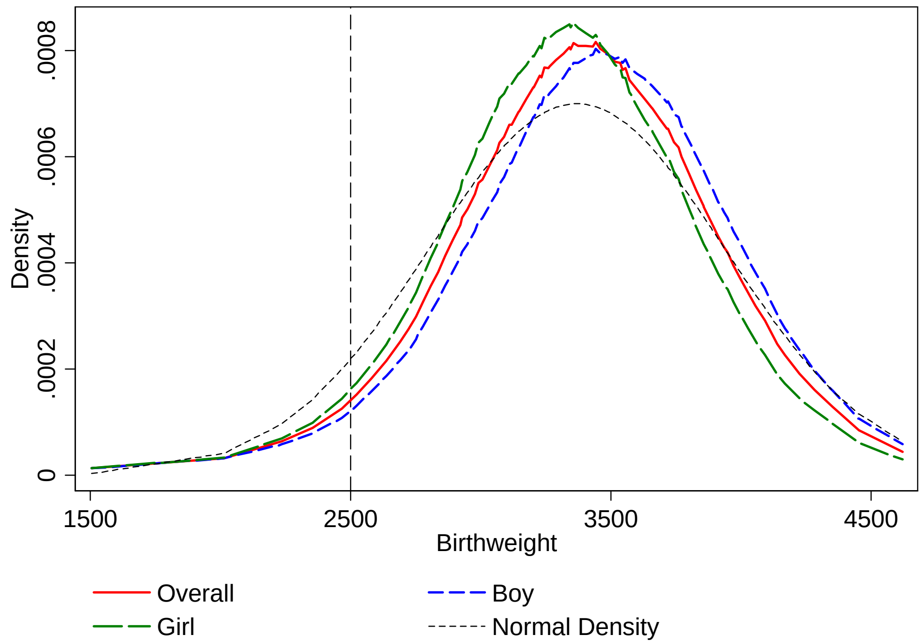

Figure A2. Probability Density Functions of Log Wages, Men 1983-85