



City Research Online

City, University of London Institutional Repository

Citation: Bera, A. K., Galvao Jr, A. F., Montes-Rojas, G. and Park, S. Y. (2010). Which quantile is the most informative? Maximum likelihood, maximum entropy and quantile regression (10/08). London, UK: Department of Economics, City University London.

This is the unspecified version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/1483/>

Link to published version: 10/08

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.



School of Social Sciences
CITY UNIVERSITY LONDON

Department of Economics

**Which Quantile is the Most Informative?
Maximum Likelihood, Maximum Entropy and
Quantile Regression**

Anil K. Bera

University of Illinois, Urbana-Champaign

Antonio F. Galvao Jr.

University of Wisconsin-Milwaukee

Gabriel Montes-Rojas*

City University London

Sung Y. Park

University of Illinois, Urbana-Champaign

Department of Economics

Discussion Paper Series

No. 10/08

* Department of Economics, City University London, Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK.

Email: Gabriel.Montes-Rojas.l@city.ac.uk

Which Quantile is the Most Informative? Maximum Likelihood, Maximum Entropy and Quantile Regression*

Anil K. Bera[†] Antonio F. Galvao Jr.[‡] Gabriel V. Montes-Rojas[§]
Sung Y. Park[¶]

Abstract

This paper studies the connections among quantile regression, the asymmetric Laplace distribution, maximum likelihood and maximum entropy. We show that the maximum likelihood problem is equivalent to the solution of a maximum entropy problem where we impose moment constraints given by the joint consideration of the mean and median. Using the resulting score functions we propose an estimator based on the joint estimating equations. This approach delivers estimates for the slope parameters together with the associated “most probable” quantile. Similarly, this method can be seen as a penalized quantile regression estimator, where the penalty is given by deviations from the median regression. We derive the asymptotic properties of this estimator by showing consistency and asymptotic normality under certain regularity conditions. Finally, we illustrate the use of the estimator with a simple application to the U.S. wage data to evaluate the effect of training on wages.

Keywords: Quantile Regression; Treatment Effects; Asymmetric Laplace Distribution

JEL classification: C14; C31

*We are very grateful to Arnold Zellner, Jushan Bai, Rong Chen, Daniel Gervini, Yongmiao Hong, Carlos Lamarche, Ehsan Soofi, Zhijie Xiao, and the participants in seminars at University of Wisconsin-Milwaukee, City University London, World Congress of the Econometric Society, Shanghai, August 2010, Latin American Meeting of the Econometric Society, Argentina, October 2009, Summer Workshop in Econometrics, Tsinghua University, Beijing, China, May 2009, South Asian and Far Eastern Meeting of the Econometrics Society, Singapore, July 2008, for helpful comments and discussions. However, we retain the responsibility for any remaining errors.

[†]Department of Economics, University of Illinois, 1407 W. Gregory Drive, Urbana, IL 61801. Tel.: +1-217-3334596; fax: +1-217-2446678. E-mail: abera@illinois.edu

[‡]Department of Economics, University of Wisconsin-Milwaukee, Bolton Hall 852, 3210 N. Maryland Ave., Milwaukee, WI 53201. E-mail: agalvao@uwm.edu

[§]Department of Economics, City University of London, 10 Northampton Square, London EC1V 0HB, U.K. E-mail: Gabriel.Montes-Rojas.1@city.ac.uk

[¶]Department of Economics, University of Illinois, 1407 W. Gregory Drive, Urbana, IL 61801, and The Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China. E-mail: sungpark@sungpark.net

1 Introduction

Different choices of loss functions determine different ways of defining the location of a random variable y . For example, squared, absolute value, and step function lead to mean, median and mode, respectively (see Manski, 1991, for a general discussion). For a given quantile $\tau \in (0, 1)$, consider the loss function in a standard quantile estimation problem,

$$L_{1,n}(\mu; \tau) = \sum_{i=1}^n \rho_{\tau}(y_i - \mu) = \sum_{i=1}^n (y_i - \mu) (\tau - 1(y_i \leq \mu)), \quad (1)$$

as proposed by Koenker and Bassett (1978). Minimizing $L_{1,n}$ with respect to the location parameter μ is identical to maximizing the likelihood based on the asymmetric Laplace probability density (ALPD):

$$f(y; \mu, \tau, \sigma) = \frac{\tau(1 - \tau)}{\sigma} \exp\left(-\frac{\rho_{\tau}(y - \mu)}{\sigma}\right), \quad (2)$$

for given τ . The well known symmetric Laplace (double exponential) distribution is a special case of (2) when $\tau=1/2$.

Several studies developed the properties of the maximum likelihood (ML) estimators based on ALPD. Hinkley and Revankar (1977) derived the asymptotic properties of the unconditional MLE under ALPD. Kotz, Kozubowski, and Podgórski (2002b) and Yu and Zhang (2005) consider alternative MLE approaches for ALPD. Moreover, models based on ALPD have been proposed in different contexts. Machado (1993) used the ALPD to derive a Schwartz information criterion for model selection for quantile regression (QR) models, and Koenker and Machado (1999) introduced a goodness-of-fit measure for QR and related inference processes. Yu and Moyeed (2001) and Geraci and Botai (2007) used a Bayesian QR approach based on the ALPD. Komunjer (2005) constructed a new class of estimators for conditional quantiles in possibly misspecified nonlinear models with time series data. The proposed estimators belong to the family of quasi-maximum likelihood estimators (QMLEs) and are based on a family of ‘tick-exponential’ densities. Under the asymmetric Laplace density, the corresponding QMLE reduces to the Koenker and Bassett (1978) linear quantile regression estimator. In addition, Komunjer (2007) developed a parametric estimator for the

risk of financial time series expected shortfall based on the asymmetric power distribution, derived the asymptotic distribution of the maximum likelihood estimator, and constructed a consistent estimator for its asymptotic covariance matrix.

Interestingly, the parameter μ in functions (1) and (2) is at the same time the location parameter, the τ -th quantile, and the mode of the ALPD. For the simple (unconditional) case, the minimization of (1) returns different order-statistics. For example, if we set $\tau = \{0.1, 0.2, \dots, 0.9\}$, the solutions are, respectively, the nine deciles of y . In order to extract important information from the data a good summary statistic would be to choose *one* order statistics accordingly the most likely value. For a symmetric distribution one would choose the median. Using the ALPD, for given τ , maximization of the corresponding likelihood function gives that particular order statistics. Thus, the main idea of this paper is to *jointly* estimate τ and the corresponding order statistic of y which can be taken as a good summary statistic of the data. The above notion can be easily extended to modeling the “conditional location” of y given covariates x , as we do in Section 2.3. In this case, the ALPD model provides a twist to the QR problem, as now τ becomes the *most likely quantile* in a regression set-up.

The aim of this paper is threefold. *First*, we show that the score functions implied by the ALPD-ML estimation are not restricted to the true data generating process being ALPD, but they arise as the solution to a maximum entropy (ME) problem where we impose moment constraints given by the joint consideration of the mean and median. By so doing, the ALPD-ML estimator combines the information in the mean and the median to capture the asymmetry of the underlying empirical distribution (see Park and Bera, 2009, for a related discussion).

Secondly, we propose a novel Z-estimator that is based on the estimating equations from the MLE score functions (which also correspond to the ME problem). We refer to this estimator as ZQR. The approximate Z-estimator do not impose that the underline distribution is ALPD. Thus, although the original motivation for using the estimating equations is based on the ALPD, the final estimator is independent of this requirement. We derive the asymptotic properties of the estimator by showing consistency and asymptotic normality under certain

regularity conditions. This approach delivers estimates for the slope parameters together with the associated ‘*most probable*’ quantile. The intuition behind this estimator works as follows. For the symmetric and unimodal case the selected quantile is the median, which coincides with the mean and mode. On the other hand, when the mean is larger than the median, the distribution is right skewed. Thus, taking into consideration the empirical distribution, there is more probability mass to the left of the distribution. As a result it is natural to consider a point estimate in a place with more probability mass. The selected τ -quantile does not necessarily lead to the mode, but to a point estimate that is most probable. This provides a new interpretation of QR and frames it within the ML and ME paradigm.

The proposed estimator has an interesting interpretation from a policy perspective. The QR analysis gives a full range of estimators that account for heterogeneity in the response variable to certain covariates. However, the proposed ZQR estimator answers the question: of all the heterogeneity in the conditional regression model, which one is more likely to be observed? In general, the entire QR process is of interest because we would like to either test global hypotheses about conditional distributions or make comparisons across different quantiles (for a discussion about inference in QR models see Koenker and Xiao, 2002). But selecting a particular quantile provides an estimator as parsimonious as ordinary least squares (OLS) or the median estimators. The proposed estimator is, therefore, a complement to the QR analysis rather than a competing alternative. This set-up also allows for an alternative interpretation of the QR analysis. Consider, for instance, the standard conditional regression set-up, $y = x'\beta + u$, and let β be partitioned into $\beta = (\beta_1, \beta_2)$. For a given value of $\beta_1 = \bar{\beta}_1$, we may be interested in finding the representative quantile of the unobservables distribution that corresponds to this level of β_1 . For such a case, instead of assuming a given quantile τ we would like to estimate it. In other words, the QR process provides us with the graph $\beta_1(\tau)$, but the graph $\tau(\beta_1)$ could be of interest too.

Finally, the *third* objective of this work is to illustrate the implementation of the proposed ZQR estimator. We apply the estimator to the estimation of quantile treatment effects of subsidized training on wages under the Job Training Partnership Act (JTPA). We discuss the relationship between OLS, median regression and ZQR estimates of the JTPA treatment

effect. We show that each estimator provides different treatment effect estimates. Moreover, we extend our ZQR estimator to Chernozhukov and Hansen (2006, 2008) instrumental variables strategy in QR.

The rest of the paper is organized as follows. Section 2 develops the ML and ME frameworks of the problem. Section 3 derives the asymptotic distribution of the estimators. In Section 4 we report a small Monte Carlo study to assess the finite sample performance of the estimator. Section 5 deals with an empirical illustration to the effect of training on wages. Finally, conclusions are in the last section.

2 Maximum Likelihood and Maximum Entropy

In this section we describe the MLE problem based on the ALPD and show its connection with the maximum entropy. We show that they are equivalent under some conditions. In the next section we will propose an Z-estimator based on the resulting estimating equations from the MLE problem, which corresponds to ME.

2.1 Maximum Likelihood

Using (2), consider the maximization of the log-likelihood function of an ALPD:

$$L_{2,n}(\mu, \tau, \sigma) = n \ln \left(\frac{1}{\sigma} \tau (1 - \tau) \right) - \sum_{i=1}^n \frac{1}{\sigma} \rho_{\tau}(y_i - \mu) = n \ln \left(\frac{1}{\sigma} \tau (1 - \tau) \right) - \frac{1}{\sigma} L_{1,n}(\mu; \tau), \quad (3)$$

with respect to μ , τ and σ . The first order conditions from (3) lead to the following estimating equations (EE):

$$\sum_{i=1}^n \frac{1}{\sigma} \left(\frac{1}{2} \text{sign}(y_i - \mu) + \tau - \frac{1}{2} \right) = 0, \quad (4)$$

$$\sum_{i=1}^n \left(\frac{1 - 2\tau}{\tau(1 - \tau)} - \frac{(y_i - \mu)}{\sigma} \right) = 0, \quad (5)$$

$$\sum_{i=1}^n \left(-\frac{1}{\sigma} + \frac{1}{\sigma^2} \rho_{\tau}(y_i - \mu) \right) = 0. \quad (6)$$

Let $(\hat{\mu}, \hat{\tau}, \hat{\sigma})$ denote the solution to this system of equations. The first equation leads to the most probable order statistic. Once we have $\hat{\tau}$, $(1 - 2\hat{\tau})$ will provide a measure of asymmetry of the distribution. Equation (6) provides a straightforward measure of dispersion, namely,

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^n \rho_{\hat{\tau}}(y_i - \hat{\mu}).$$

Then, the loss function corresponding to (3) can be rewritten as a two-parameter loss function

$$-\frac{1}{n} L_{2,n}(\mu, \tau) = \ln \left(\frac{1}{n} L_{1,n}(\mu; \tau) \right) - \ln(\tau(1 - \tau)). \quad (7)$$

This determines that $L_{2,n}(\mu, \tau, \sigma)$ can be seen as a penalized quantile optimization function, where we minimize $\ln \left(\frac{1}{n} L_{1,n}(\mu; \tau) \right)$ and penalize it by $-\ln(\tau(1 - \tau))$. The penalty can be interpreted as the cost of deviating from the median, i.e. for $\tau = 1/2$, $-\ln(\tau(1 - \tau)) = -\ln(1/4)$ is the minimum, while for either $\tau \rightarrow 0$ or $\tau \rightarrow 1$ the penalty goes to $+\infty$.

It is important to note that the structure of the estimating functions suggests that the solution to the MLE problem can be obtained by first obtaining every quantile of the distribution, and then plugging them (with the corresponding estimator for σ) in (5) until this equation is satisfied (if the solution is unique). In other words, given all the quantiles of y , the problem above selects the most likely quantile as if the distribution of y were ALPD.

2.2 Maximum Entropy

The ALPD can be characterized as a maximum entropy density obtained by maximizing Shannon's entropy measure subject to two moment constraints (see Kotz, Kozubowski, and Podgórsk, 2002a):

$$f_{ME}(y) \equiv \arg \max_f - \int f(y) \ln f(y) dy \quad (8)$$

subject to

$$E|y - \mu| = c_1, \quad (9)$$

$$E(y - \mu) = c_2, \quad (10)$$

and the normalization constraint, $\int f(y)dy = 1$, where c_1 and c_2 are known constants. The solution to the above optimization problem using the Lagrangian has the familiar exponential form

$$f_{ME}(y : \mu, \lambda_1, \lambda_2) = \frac{1}{\Omega(\theta)} \exp[-\lambda_1|y - \mu| - \lambda_2(y - \mu)], \quad -\infty < y < \infty, \quad (11)$$

where λ_1 and λ_2 are the Lagrange multipliers corresponding to the constraints (9) and (10), respectively, $\theta = (\mu, \lambda_1, \lambda_2)'$ and $\Omega(\theta)$ is the normalizing constant. Note that $\lambda_1 \in \mathbb{R}^+$ and $\lambda_2 \in [-\lambda_1, \lambda_1]$ so that $f_{ME}(y)$ is well-defined. Symmetric Laplace density (LD) is a special case of ALPD when λ_2 is equal to zero.

Interestingly, the constraints (9) and (10) capture, respectively, the dispersion and asymmetry of the ALPD. The marginal contribution of (10) is measured by the Lagrangian multiplier λ_2 . If λ_2 is close to 0, then (10) does not have useful information for the data, and therefore, the symmetric LD is the most appropriate. In this case, μ is known to be the median of the distribution. On the other hand, when λ_2 is not close to zero, it measures the degree of asymmetry of the ME distribution. Thus the non-zero value of λ_2 makes $f_{ME}(\cdot)$ deviate from the symmetric LD, and therefore, changes the location, μ , of the distribution to adhere the maximum value of the entropy (for general notion of entropy see Soofi and Retzer, 2002).

Let us write (9) and (10), respectively, as

$$\int \phi_1(y, \mu) f_{ME}(y : \mu, \lambda_1, \lambda_2) dy = 0 \quad \text{and} \quad \int \phi_2(y, \mu) f_{ME}(y : \mu, \lambda_1, \lambda_2) dy = 0,$$

where $\phi_1(y, \mu) = |y - \mu| - c_1$ and $\phi_2(y, \mu) = (y - \mu) - c_2$. By substituting the solution $f_{ME}(y : \mu, \lambda_1, \lambda_2)$ into the Lagrangian of the maximization problem in (8), we obtain the profiled objective function

$$h(\lambda_1, \lambda_2, \mu) = \ln \int \exp \left[- \sum_{j=1}^2 \lambda_j \phi_j(y, \mu) \right] dy. \quad (12)$$

The parameters λ_1 , λ_2 and μ can be estimated by solving the following saddle point problem (Kitamura and Stutzer, 1997)

$$\hat{\mu}_{ME} = \arg \max_{\mu} \ln \int \exp \left[- \sum_{j=1}^2 \hat{\lambda}_{j,ME} \phi_j(y, \mu) \right] dy,$$

where $\hat{\lambda}_{ME} = (\hat{\lambda}_{1,ME}, \hat{\lambda}_{2,ME})$ is given by

$$\hat{\lambda}_{ME}(\mu) = \arg \min_{\lambda} \ln \int \exp \left[- \sum_{j=1}^2 \lambda_j \phi_j(y, \mu) \right] dy.$$

Solving the above saddle point problem is relatively easy since the profiled objective function has the exponential form. However, generally, c_1 and c_2 are not known or functions of parameters and Lagrange multipliers in a non-linear fashion. Moreover, in some cases, the closed form of c_1 and c_2 is not known. In order to deal with this problem, we simply consider the sample counterpart of the moments c_1 and c_2 , say, $c_1 = (1/n) \sum_{i=1}^n |y_i - \mu|$ and $c_2 = (1/n) \sum_{i=1}^n (y_i - \mu)$. Then, it can be easily shown that the profiled objective function is simply the negative log-likelihood function of asymmetric Laplace density, i.e., $h(\lambda_1, \lambda_2, \mu) = -(1/n) L_{2,n}(\mu, \tau, \sigma)$ (see Ebrahimi, Soofi, and Soyer, 2008). In this case, $\hat{\mu}_{ME}$ and $\hat{\lambda}_{ME}$ satisfy the following first order conditions $\partial h / \partial \mu = 0$, $\partial h / \partial \lambda_2 = 0$ and $\partial h / \partial \lambda_1 = 0$, respectively:

$$-\frac{\lambda_1}{n} \sum_{i=1}^n \text{sign}(y_i - \mu) - \lambda_2 = 0. \quad (13)$$

$$\frac{2\lambda_2}{(\lambda_1 + \lambda_2)(\lambda_1 - \lambda_2)} + \frac{1}{n} \sum_{i=1}^n (y_i - \mu) = 0, \quad (14)$$

$$-\frac{1}{\lambda_1} \frac{\lambda_1^2 + \lambda_2^2}{\lambda_1^2 - \lambda_2^2} + \frac{1}{n} \sum_{i=1}^n |y_i - \mu| = 0, \quad (15)$$

Equations (13)-(15) are a re-parameterized version of (4)-(6). In fact, from a comparison of (2) and (11) we can easily see that $\lambda_1 = 1/(2\sigma)$, $\lambda_2 = (2\tau - 1)/(2\sigma)$ and $\Omega(\theta) = \sigma/(\tau(1 - \tau))$. Given λ_1 the degree of asymmetry is explained by λ_2 that is proportionally equal to $2\tau - 1$ in ALPD. Note that $\lambda_2 = 0$ when $\tau = 0.5$, i.e., μ is the median. Thus finding the most appropriate degree of asymmetry is equivalent to estimating τ based on the ML method.

The role of the two moment constraints can be explained by the linear combination of two moment functions, $|y - \mu|$ and $(y - \mu)$. Figure 1 plots $g(y; \lambda_1, \lambda_2, \mu) = \lambda_1|y - \mu| + \lambda_2(y - \mu)$ with three different values of λ_2 , $\lambda_1 = 1$, and $\mu = 0$. In general, $g(y; \lambda_1, \lambda_2, \mu)$ can be seen as a loss function. Clearly, this loss function is symmetric when $\lambda_2 = 0$. When $\lambda_2 = 1/3$, $g(\cdot)$ is tilted so that it puts more weight on the positive values in order to attain the maximum of the Shannon's entropy (and the reverse is true for $\lambda_2 = -1/3$). This naturally yields the asymmetric behavior of the resulting ME density.

[Figure 1]

2.3 Linear Regression Model

Now consider the conditional version of the above, by taking a linear model of the form $y = x'\beta + u$, where the parameter of interest is $\beta \in \mathbb{R}^p$, x refers to a p -vector of exogenous covariates, and u denotes the unobservable component in the linear model. As noted in Angrist, Chernozhukov, and Fernández-Val (2006), QR provides the best linear predictor for y under the asymmetric loss function

$$L_{3,n}(\beta; \tau) = \sum_{i=1}^n \rho_{\tau}(y_i - x'_i\beta) = \sum_{i=1}^n ((y_i - x'_i\beta) (\tau - 1(y_i \leq x'_i\beta))), \quad (16)$$

where β is assumed to be a function of the *fixed* quantile τ of the unobservable components, that is $\beta(\tau)$. If u is assumed to follow an ALPD, the log-likelihood function is

$$\begin{aligned} L_{4,n}(\beta, \tau, \sigma) &= n \ln \left(\frac{1}{\sigma} \tau (1 - \tau) \right) - \sum_{i=1}^n \left(\frac{1}{\sigma} \rho_{\tau}(y_i - x'_i\beta) \right) \\ &= n \ln \left(\frac{1}{\sigma} \tau (1 - \tau) \right) - \frac{1}{\sigma} L_{3,n}(\beta; \tau). \end{aligned} \quad (17)$$

Estimating β in this framework provides the marginal effect of x on the τ -quantile of the conditional quantile function of y .

Computationally, the MLE can be obtained by simulating a grid of quantiles and choosing the quantile that maximizes (17), or by solving the estimating equations, $\nabla L_{4,n}(\beta, \tau, \sigma) = 0$,

i.e.,

$$\frac{\partial L_{4,n}(\beta, \tau, \sigma)}{\partial \beta} = \sum_{i=1}^n \frac{1}{\sigma} \left(\frac{1}{2} \text{sign}(y_i - x_i' \beta) + \tau - \frac{1}{2} \right) x_i = 0, \quad (18)$$

$$\frac{\partial L_{4,n}(\beta, \tau, \sigma)}{\partial \tau} = \sum_{i=1}^n \left(\frac{1 - 2\tau}{\tau(1-\tau)} - \frac{(y_i - x_i' \beta)}{\sigma} \right) = 0, \quad (19)$$

$$\frac{\partial L_{4,n}(\beta, \tau, \sigma)}{\partial \sigma} = \sum_{i=1}^n \left(-\frac{1}{\sigma} + \frac{1}{\sigma^2} \rho_{\tau}(y_i - x_i' \beta) \right) = 0. \quad (20)$$

As we stated before, $L_{4,n}$ can be written as a penalized QR problem loss function that depends only on (β, τ) :

$$-\frac{1}{n} L_{4,n}(\beta, \tau) = \ln \left(\frac{1}{n} L_{3,n}(\beta; \tau) \right) - \ln(\tau(1-\tau)), \quad (21)$$

and the interpretation is the same as discussed in section 2.1.

3 A Z-estimator for Quantile Regression

In this section we propose a Z-estimator based on the score functions from equations (18)-(20). Thus, although the original motivation for using the estimating equations is based on the ALPD, the final estimator is independent of this requirement. Let $\|\cdot\|$ be the Euclidean norm and $\theta = (\beta, \tau, \sigma)'$. Moreover, define the estimating functions

$$\psi_{\theta}(y, x) = \begin{pmatrix} \psi_{1\theta}(y, x) \\ \psi_{2\theta}(y, x) \\ \psi_{3\theta}(y, x) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma} (\tau - 1(y < x' \beta)) x \\ \frac{1-2\tau}{\tau(1-\tau)} - \frac{(y-x'\beta)}{\sigma} \\ -\frac{1}{\sigma} + \frac{1}{\sigma^2} \rho_{\tau}(y - x' \beta) \end{pmatrix},$$

and the estimating equations

$$\Psi_n(\theta) = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \frac{1}{\sigma} (\tau - 1(y_i < x_i' \beta)) x_i \\ \frac{1-2\tau}{\tau(1-\tau)} - \frac{(y_i-x_i'\beta)}{\sigma} \\ -\frac{1}{\sigma} + \frac{1}{\sigma^2} \rho_{\tau}(y_i - x_i' \beta) \end{pmatrix} = \frac{1}{n} \sum_{i=1}^n \psi_{\theta}(y_i, x_i) = 0.$$

A Z-estimator $\hat{\theta}_n$ is the approximate zero of the above data-dependent function that satisfies $\|\Psi_n(\hat{\theta}_n)\| \xrightarrow{p} 0$.

The implementation of the estimator is simple. As discussed in the previous section, an iteration algorithm can be used to solve for the estimates in the estimating equations above. Computationally, the estimates can be obtained by constructing a grid for quantiles τ and solving the QR problem as in (18) and (19) to find $\hat{\beta}(\tau)$ and $\hat{\sigma}(\tau)$. Finally, we estimate the quantile $\hat{\tau}$ that finds an approximate zero in (20). This algorithm is similar to the one proposed in Hinkley and Revankar (1977) and Yu and Zhang (2005) that compute the estimators for MLE under the ALPD. We find that the algorithm converges fast and is very precise.

In the proposed Z-estimator the interpretation of the parameter β is analogous to the interpretation of the location parameter in the QR literature. As in the least squares case, the scale parameter σ can be interpreted as the expected value of the loss function, which in the QR case corresponds to the expectation of the $\rho_\tau(\cdot)$ function. Finally, τ captures a measure of asymmetry of the underline distribution of $y|x$ and also is associated with the most probable quantile. In the Appendix A we discuss the interpretation of these parameters in more detail.

We introduce the following assumptions to derive the asymptotic properties.

Assumption 1. Let $y_i = x_i'\beta_0 + u_i$, $i = 1, 2, \dots, n$, where (y_i, x_i) is independent and identically distributed (i.i.d.), and x_i is independent of $u_i, \forall i$.

Assumption 2. The conditional distribution function of y , $G(y|x)$, is absolutely continuous with conditional densities, $g(y|x)$, with $0 < g(\cdot|\cdot) < \infty$.

Assumption 3. Let Θ be a compact set, with $\theta = (\beta, \tau, \sigma)' \in \Theta$, where $\beta \in \mathcal{B} \subset \mathbb{R}^p$, $\tau \in \mathcal{T} \subset (0, 1)$, and $\sigma \in \mathcal{S} \subset \mathbb{R}^+$;

Assumption 4. $E\|x\|^{2+\epsilon} < \infty$, and $E\|y\|^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

Assumption 5. (i) Define $\Psi(\theta) = E[\psi_\theta(y, x)]$. Assume that $\Psi(\theta_0) = 0$ for a unique $\theta_0 \in \text{int}\Theta$. (ii) Define $\Psi_n(\theta) = \mathbb{E}_n[\psi_\theta(y, x)] = \frac{1}{n} \sum_{i=1}^n \psi_\theta(y_i, x_i)$. Assume that $\|\Psi_n(\hat{\theta}_n)\| = o_p(n^{-1/2})$.

Assumption A1 considers the usual linear model and imposes i.i.d. to facilitate the proofs.

Assumption A2 is common in the QR literature and restricts the conditional distribution of the dependent variable. Assumption A3 imposes compactness of the parameter space, and A4 is important to guarantee the asymptotic behavior of the estimator. The first part of A4 is usual in QR literature and second part in least squares literature. Finally Assumption A5 imposes an identifiability condition and ensure that the solution to the estimating equations is “nearly-zero”, and it deserves further discussion.

The first part of A5 imposes a unique solution condition. Similar restrictions are frequently used in the QR literature to satisfy $E[\psi_{1\theta}(y, x)] = 0$ for a unique β and any given τ . This condition also appears in the M and Z estimators literatures. Uniqueness in QR is a very delicate subject and is actually imposed. For instance, Chernozhukov, Fernández-Val, and Melly (2009, p. 49) propose an approximate Z-estimator for QR process and assume that the true parameter $\beta_0(\tau)$ solves $E[(\tau - 1\{y \leq X'\beta_0(\tau)\})X] = 0$. Angrist, Chernozhukov, and Fernández-Val (2006) impose a uniqueness assumption of the form: $\beta(\tau) = \arg \min_{\beta} E[\rho_{\tau}(y - x'\beta)]$ is unique (see for instance their Theorems 1 and 2). See also He and Shao (2000) and Schennach (2008) for related discussion.

It is possible to impose more primitive conditions to ensure uniqueness. These conditions are explored and discussed in Theorem 2.1 in Koenker (2005, p. 36). If the y 's have a bounded density with respect to Lebesgue measure then the observations (y, x) will be in general position with probability one and a solution exists.¹ However, uniqueness cannot be ensured if the covariates are discrete (e.g. dummy variables). If the x 's have a component that have a density with respect to a Lebesgue measure, then multiple optimal solutions occur with probability zero and the solution is unique. However, these conditions are not very attractive, and uniqueness is in general imposed as an assumption.

Note that the usual assumptions of uniqueness in QR described above for $E[\psi_{1\theta}(y, x)] = 0$ guarantee that β is unique for any given τ . Combining these assumptions and bounded moments we guarantee uniqueness for $E[\psi_{3\theta}(y, x)] = 0$, because $\sigma = E[\rho_{\tau}(y - x\beta)]$ such that for each τ , σ is unique. With respect to the second equation $E[\psi_{2\theta}(y, x)] = 0$, it is satisfied

¹See definition 2.1 in Koenker (2005) for a definition of general position.

if

$$\frac{1-2\tau}{\tau(1-\tau)} = \frac{E[y - x'\beta]}{\sigma}.$$

Therefore, for unique β and σ , the right hand side of the above equation is unique. Since $\frac{1-2\tau}{\tau(1-\tau)}$ is a continuous and strictly decreasing, τ is also unique.²

The second part of A5 is used to ensure that the solution to the approximated working estimating equations is close to zero. The solution for the estimating equations, $\Psi_n(\hat{\theta}_n) = 0$, does not hold in general. In most cases, this condition is actually equal to zero, but least absolute deviation of linear regression is one important exception. The indicator function in the first estimating equations determines that it may not have an exact zero. It is common in the literature to work with M and Z estimators $\hat{\theta}_n$ of θ_0 that satisfy $\sum_{i=1}^n \psi(x_i, \hat{\theta}_n) = o_p(\delta_n)$, for some sequence δ_n . For example, Huber (1967) considered $\delta_n = \sqrt{n}$ for asymptotic normality, and Hinkley and Revankar (1977) verified the condition for the unconditional asymmetric double exponential case. This condition also appears in the quantile regression literature, see for instance He and Shao (1996) and Wei and Carroll (2009). In addition, in the approximate Z-estimator for quantile process in Chernozhukov, Fernández-Val, and Melly (2009), they have that the empirical moment functions $\hat{\Psi}(\theta, u) = E_n[g(W_i, \theta, u)]$, for each $u \in T$, the estimator $\hat{\theta}(u)$ satisfies $\|\hat{\Psi}(\hat{\theta}(u), u)\| \leq \inf_{\theta \in \Theta} \|\hat{\Psi}(\theta, u)\| + \epsilon_n$ where $\epsilon_n = o(n^{-1/2})$. For the quantile regression case, Koenker (2005, p. 36) comments that the absence of a zero to the problem $\Psi_{1n}(\hat{\beta}_n(\tau)) = 0$, where $\hat{\beta}_n(\tau)$ is the quantile regression optimal solution for a given τ and σ , “is unusual, unless the y_i ’s are discrete.” Here we follow the standard conditions for M and Z estimators and impose A5(ii). For a more general discussion about this condition on M and Z estimators see e.g. Kosorok (2008, pp. 399-407).

Now we move our attention to the asymptotic properties of the estimator.

Theorem 1 *Under Assumptions A1-A5, $\|\hat{\theta}_n - \theta_0\| \xrightarrow{p} 0$.*

Proof: In order to show consistency we check the conditions of Theorem 5.9 in van der

²Note that $m(\tau) \equiv \frac{1-2\tau}{\tau(1-\tau)}$ is a continuous function, has a unique zero at $\tau = 1/2$ and $m(\tau) > 0$ for $\tau < 1/2$, $m(\tau) < 0$ for $\tau > 1/2$. As $\tau \rightarrow 0$, $m(\tau) \rightarrow +\infty$, and as $\tau \rightarrow 1$, $m(\tau) \rightarrow -\infty$. Finally, $\frac{dm(\tau)}{d\tau} = \frac{-2\tau(1-\tau) - (1-2\tau)^2}{\tau^2(1-\tau)^2} = \frac{-2\tau + 2\tau^2 - 1 + 4\tau - 4\tau^2}{\tau^2(1-\tau)^2} = \frac{-1 + 2\tau - 2\tau^2}{\tau^2(1-\tau)^2} = \frac{-1 + 2\tau(1-\tau)}{\tau^2(1-\tau)^2} < 0$ for any $\tau \in (0, 1)$.

Vaart (1998). Define $\mathcal{F} \equiv \{\psi_\theta(y, x), \theta \in \Theta\}$, and recall that $\Psi_n(\theta) = \frac{1}{n} \sum_{i=1}^n \psi_\theta(y, x)$ and $\Psi(\theta) = E[\psi_\theta(y, x)]$. First note that, under conditions A3 and A5, the function $\Psi(\theta)$ satisfies,

$$\inf_{\theta: d(\theta, \theta_0) \geq \epsilon} \|\Psi(\theta)\| > 0 = \|\Psi(\theta_0)\|,$$

because for a compact set Θ and a continuous function Ψ , uniqueness of θ_0 as a zero implies this condition (see van der Vaart, 1998, p.46).

Now we need to show that $\sup_{\theta \in \Theta} \|\Psi_n(\theta) - \Psi(\theta)\| \xrightarrow{p} 0$. By Lemma A1 in the Appendix B we know that the class \mathcal{F} is Donsker. Donskerness implies a uniform law of large numbers such that

$$\sup_{\theta \in \Theta} |\mathbb{E}_n[\psi_\theta(y, x)] - E[\psi_\theta(y, x)]| \xrightarrow{p} 0,$$

where $f \mapsto \mathbb{E}_n[f(w)] = \frac{1}{n} \sum_{i=1}^n f(w_i)$. Hence we have $\sup_{\theta \in \Theta} \|\Psi_n(\theta) - \Psi(\theta)\| \xrightarrow{p} 0$.

Finally, from assumptions A1-A5 the problem has a unique root and also we have $\|\Psi_n(\hat{\theta}_n)\| \xrightarrow{p} 0$. Thus, all the conditions in Theorem 5.9 of van der Vaart (1998) are satisfied and $\|\hat{\theta}_n - \theta_0\| \xrightarrow{p} 0$. ■

After showing consistency we move our attention to the asymptotic normality of the estimator. In order to derive the limiting distribution define

$$V_{1\theta} = E[\psi_\theta(y, x)\psi_\theta(y, x)'], \quad (22)$$

$$V_{2\theta} = \frac{\partial E[\psi_\theta(y, x)]}{\partial \theta'}. \quad (23)$$

Here,

$$V_{1\theta} = \begin{bmatrix} \frac{1}{\sigma} \tau(1-\tau) E[xx'] & \frac{E[((1-2\tau)\text{sign}(y-x'\beta) - (1-2\tau)^2)x']}{2\sigma\tau(1-\tau)} - E\left[\frac{1}{\sigma^2} \rho_\tau(y-x'\beta)x'\right] & \frac{1}{2\sigma^3} E[\rho_\tau(y-x'\beta) (\text{sign}(y-x'\beta) - (1-2\tau))x'] \\ \cdot & \frac{(1-2\tau)^2}{\tau^2(1-\tau)^2} + E\left[\frac{1}{\sigma^2} (y-x'\beta)^2\right] - 2\frac{(1-2\tau)}{\tau(1-\tau)} E\left[\frac{1}{\sigma} (y-x'\beta)\right] & \frac{1}{\sigma^2} E[\rho_\tau(y-x'\beta) \left(\frac{(1-2\tau)}{\tau(1-\tau)} - \frac{1}{\sigma} (y-x'\beta)\right)] \\ \cdot & \cdot & \frac{1}{\sigma^4} E[\rho_\tau^2(y-x'\beta)] + \frac{1}{\sigma^2} - \frac{1}{\sigma^3} E[\rho_\tau(y-x'\beta)] \end{bmatrix}$$

and

$$V_{2\theta} = \begin{bmatrix} -\frac{E[g(x'\beta|x)xx']}{\sigma} & \frac{1}{\sigma} E[x] & 0 \\ \cdot & \frac{-1+2\tau-2\tau^2}{\tau^2(1-\tau)^2} & \frac{1}{\sigma^2} E[(y-x'\beta)] \\ \cdot & \cdot & -\frac{1}{\sigma^2} \end{bmatrix}.$$

Note that when $y|x \sim ALPD(x'\beta, \tau, \sigma)$, then $V_{1\theta} = V_{2\theta}$.

Assumption 6. Assume that $V_{1\theta_0}$ and $V_{2\theta_0}$ exist and are finite, and $V_{2\theta_0}$ is invertible.

Chernozhukov, Fernández-Val, and Melly (2009) calculated equations (22) and (23) in the quantile process as an approximate Z-estimator.

Now we state the asymptotic normality result.

Theorem 2 *Under Assumptions 1-6,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow N(0, V_{2\theta_0}^{-1}V_{1\theta_0}V_{2\theta_0}^{-1}).$$

Proof: First, combining Theorem 1 and second part of Lemma A1, we have

$$\mathbb{G}_n\psi_{\hat{\theta}_n}(y, x) = \mathbb{G}_n\psi_{\theta_0}(y, x) + o_p(1),$$

where $f \mapsto \mathbb{G}_n[f(w)] = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(w_i) - Ef(w_i))$. Rewriting we have

$$\sqrt{n}\mathbb{E}_n\psi_{\hat{\theta}_n}(y, x) = \sqrt{n}E\psi_{\hat{\theta}_n}(y, x) + \mathbb{G}_n\psi_{\theta_0}(y, x) + o_p(1). \quad (24)$$

By assumption A5

$$\|\mathbb{E}_n\psi_{\hat{\theta}_n}(y, x)\| = o_p(n^{-1/2}) \quad \text{and} \quad E[\psi_{\theta_0}(y, x)] = 0.$$

Now consider the first element of the right hand side of (24). By a Taylor expansion about $\hat{\theta}_n = \theta_0$ we obtain

$$E[\psi_{\hat{\theta}_n}(y, x)] = E[\psi_{\theta_0}(y, x)] + \frac{\partial E[\psi_{\theta}(y, x)]}{\partial \theta'} \Big|_{\theta=\theta_0} (\hat{\theta}_n - \theta_0) + o_p(1), \quad (25)$$

where

$$\frac{\partial E[\psi_{\theta}(y, x)]}{\partial \theta'} \Big|_{\theta=\theta_0} = \frac{\partial}{\partial \theta'} E \left(\begin{array}{c} \frac{1}{\sigma} (\tau - 1(y < x'\beta)) x \\ \left(\frac{1-2\tau}{\tau(1-\tau)} - \frac{(y-x'\beta)}{\sigma} \right) \\ -\frac{1}{\sigma} + \frac{1}{\sigma^2} \rho_{\tau}(y - x'\beta) \end{array} \right) \Big|_{\theta=\theta_0}.$$

Since by condition A6, $\frac{\partial E[\psi_{\theta}(y, x)]}{\partial \theta'} \Big|_{\theta=\theta_0} = V_{2\theta_0}$, equation (25) can be rewritten as

$$E[\psi_{\theta}(y, x)] \Big|_{\theta=\hat{\theta}_n} = V_{2\theta_0}(\hat{\theta}_n - \theta_0) + o_p(1). \quad (26)$$

Using Assumption A5 (ii), from (24) we have

$$o_p(1) = \sqrt{n}E\psi_{\hat{\theta}_n}(y, x) + \mathbb{G}_n\psi_{\theta_0}(y, x) + o_p(1),$$

and using the above approximation given in (26)

$$o_p(1) = V_{2\theta_0}\sqrt{n}(\hat{\theta}_n - \theta_0) + \mathbb{G}_n\psi_{\theta_0}(y, x) + o_p(1).$$

By invertibility of $V_{2\theta_0}$ in A6,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V_{2\theta_0}^{-1}\mathbb{G}_n\psi_{\theta_0}(y, x) + o_p(1). \quad (27)$$

Finally, from Lemma A1 $\theta \mapsto \mathbb{G}_n\psi_{\theta}(y, x)$ is stochastic equicontinuous. So, stochastic equicontinuity and ordinary CLT imply that $\mathbb{G}_n\psi_{\theta}(y, x) \Rightarrow z(\cdot)$ converges to a Gaussian process with variance-covariance function defined by

$$V_{1\theta_0} = E[\psi_{\theta}(y, x)\psi_{\theta}(y, x)'] \Big|_{\theta=\theta_0}. \text{ Therefore, from (27)}$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow V_{2\theta_0}^{-1}z(\cdot),$$

so that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow N(0, V_{2\theta_0}^{-1}V_{1\theta_0}V_{2\theta_0}^{-1}).$$

■

4 Monte Carlo Simulations

In this section we provide a glimpse into the finite sample behavior of the proposed ZQR estimator. Two simple versions of our basic model are considered in the simulation experiments. In the first, reported in Table 1, the scalar covariate, x_i , exerts a pure location shift effect. In the second, reported in Table 2, x_i has both a location and scale shift effects. In the former case the response, y_i , is generated by the model,

$$y_i = \alpha + \beta x_i + u_i,$$

while in the latter case,

$$y_i = \alpha + \beta x_i + (1 + \gamma)u_i,$$

where u_i are i.i.d. innovations generated according to a standard normal distribution, t_3 distribution, χ_3^2 centered at the mean, Laplace distribution (i.e. $\tau = 0.5$), and ALPD with $\tau = 0.25$.³ In the location shift model x_i follows a standard normal distribution; in the location-scale shift model, it follows a χ_3^2 . We set $\alpha = \beta = 1$ and $\gamma = 0.5$. Our interest is on the effect of the covariates in terms of bias and root mean squared error (RMSE). We carry out all the experiments with sample size $n = 200$ and 5,000 replications. Three estimators are considered: our proposed ZQR estimator, quantile regression at the median (QR), and ordinary least squares (OLS). We pay special attention to the estimated quantile $\hat{\tau}$ in the ZQR.

Table 1 reports the results for the location shift model. In all cases we compute the bias and RMSE with respect to $\beta = 1$. Bias is close to zero in all cases. In the Gaussian setting, as expected, we observe efficiency loss in ZQR and QR estimates compared to that of OLS. Under symmetric distributions, normal, t_3 , and Laplace, the estimated quantile of interest $\hat{\tau}$ in the ZQR is remarkably close to 0.5. In the χ_3^2 case, the ZQR estimator performs better than the QR and OLS procedures. Note that the estimated quantile for the χ_3^2 is 0.081, consistent with the fact that the underline distribution is right skewed. Finally, for the ALPD(0.25) case, ZQR produces the estimated quantile ($\hat{\tau} = 0.248$) rightly close to 0.25, and also has a smaller RMSE. Overall, Table 1 shows that the ZQR estimator retains the robustness properties of the QR estimator, although we do not specify a particular quantile of interest.

[Table 1]

In the location-scale version of the model we adopt the same distributions for generating the data. For this case the effect of the covariate x_i on quantile of interest response in QR is given by $\beta(\tau) = \beta + \gamma Q_u(\tau)$. In ZQR we compute bias and RMSE by averaging estimated τ from 5,000 replications. The results are summarized in Table 2. The results for the normal,

³Although not reported, similar results were obtained for ALPD with $\tau = 0.75$.

t_3 and Laplace distributions are similar to those in the location model, showing that all point estimates are approximately unbiased. As expected, OLS outperforms ZQR and QR in the normal case, but the opposite occurs in the t_3 and Laplace distributions. In the χ_3^2 case, the estimated quantile is $\hat{\tau} = 0.086$. For the ALPD(0.25) distribution, the best performance is obtained for the ZQR estimator.

[Table 2]

5 Empirical Illustration: The Effect of Job Training on Wages

The effect of policy variables on distributional outcomes are of fundamental interest in empirical economics. Of particular interest is the estimation of the quantile treatment effects (QTE), that is, the effect of some policy variable of interest on the different quantiles of a conditional response variable. Our proposed estimator complements the QTE analysis by providing a parsimonious estimator at the most probable quantile value.

We apply the estimator to the study of the effect of public-sponsored training programs. As argued in LaLonde (1995), public programs of training and employment are designed to improve participant’s productive skills, which in turn would affect their earnings and dependency on social welfare benefits. We use the Job Training Partnership Act (JTPA), a public training program that has been extensively studied in the literature. For example, see Bloom, Orr, Bell, Cave, Doolittle, Lin, and Bos (1997) for a description, and Abadie, Angrist, and Imbens (2002) for QTE analysis. The JTPA was a large publicly-funded training program that began funding in October 1983 and continued until late 1990’s. We focus on the Title II subprogram, which was offered only to individuals with “barriers to employment” (long-term use of welfare, being a high-school drop-out, 15 or more recent weeks of unemployment, limited English proficiency, physical or mental disability, reading proficiency below 7th grade level or an arrest record). Individuals in the randomly assigned JTPA treatment group were offered training, while those in the control group were excluded

for a period of 18 months. Our interest lies in measuring the effect of a training offer and actual training on of participants' future earnings.

We use the database in Abadie, Angrist, and Imbens (2002) that contains information about adult male and female JTPA participants and non-participants. Let z denote the indicator variable for those receiving a JTPA offer. Of those offered, 60% did training; of those in the control group, less than 2% did training. For our purposes of illustrating the use of ZQR, we first study the effect of receiving a JTPA offer on log wages, and later we pursue instrumental variables estimation in the ZQR context. Following Abadie, Angrist, and Imbens (2002) we use a linear regression specification model, where the JTPA offer enters in the equation as a dummy variable.⁴ We consider the following regression model:

$$y = z\gamma + x\beta + u,$$

where the dependent variable y is the logarithm of 30 month accumulated earnings (we exclude individuals without earnings), z is a dummy variable for the JTPA offer, x is a set of exogenous covariates containing individual characteristics, and u is an unobservable component. The parameter of interest is γ that provides the effect of the JTPA training offer on wages.

[Table 3 and Figures 2, and 3]

First, we compute the QR process for all $\tau \in (0.05, 0.95)$ and the results are presented in Figure 2. The JTPA effect estimates for QR and OLS appear in Table 3. Interestingly, with exception of low quantiles, the effect of JTPA is decreasing in τ , which implies that those individuals in the high quantiles of the conditional wage distribution benefited less from the JTPA training. Second, by solving equation (19) we obtain that the most probable quantile $\hat{\tau} = 0.84$. This is further illustrated in Figure 3, and this means that the distribution of unobservables is negatively skewed. This value is denoted by a vertical solid line, together with the 95% confidence interval given by the vertical parallel dotted lines. From Table 3 we

⁴Linear regression models are common in the QTE literature to accomodate several control variables capturing individual characteristics. See for instance Chernozhukov and Hansen (2006, 2008) and Firpo (2007).

observe that the training effect estimate from mean and median regressions are, respectively, 0.075 (0.032) and 0.100 (0.033) which are similar, however they both are larger than the ZQR estimate of 0.045 (0.022).⁵ Figure 2 shows that QR estimates in the upper tail of the distribution have smaller standard errors, which suggests that by choosing the most likely quantile the ZQR procedure implicitly solves for the smallest standard error QR estimator. The results show that for the most probable quantile, $\hat{\tau} = 0.84$ (0.051), the effect of training is different from the mean and median effects. From a policy maker perspective, if one is asked to report the effect of training on wage, it could be done through the mean effect (0.075), the median effect (0.100) or even the entire conditional quantile function as in Figure 2; our analysis recommends reporting the most likely effect (0.045) coming from the most probable quantile $\hat{\tau} = 0.84$. Using the above model, the fit of the data reveals that the upper quantiles are informative, and the ZQR estimator is appropriate to describe the effect of JTPA on earnings.

As argued in the Introduction, the ZQR framework allows for a different interpretation of the QR analysis. Suppose that we are interested in a targeted treatment effect of $\bar{\gamma} = 0.1$, and we would like to get the representative quantile of the unobservables distribution that will most likely have this effect. This corresponds to estimating the ZQR parameters for $y - z\bar{\gamma} = x\beta + u$. In this case, we obtain an estimated most likely quantile of $\widehat{\tau(\bar{\gamma})} = 0.85$.

To value the option of treatment is an interesting exercise in itself, but policy makers may be more interested in the effect of actual training rather than the possibility of training. In this case the model of interest is

$$y = d\alpha + x\beta + u$$

where d is a dummy variable indicating if the individual actually completed the JTPA training. We have strong reasons to believe that $cov(d, u) \neq 0$ and therefore OLS and QR estimates will be biased. In this case, while the JTPA offer is random, those individuals who decide to undertake training do not constitute a random sample of the population. Rather, they are likely to be more motivated individuals or those that value training the

⁵The numbers in parenthesis are the corresponding standard errors.

most. However, the exact nature of this bias is unknown in terms of quantiles. Figure 4 reports the entire quantile process and OLS for the above equation. Interestingly the effect of training on wages is monotonically decreasing in τ . The selection of the most likely quantile determines that as in the previous case $\hat{\tau} = 0.84$.

[Figure 4]

In order to solve for the potential endogeneity, and following Abadie, Angrist, and Imbens (2002), z can be used as a valid instrument for d . The reason is that it is exogenous as it was a randomized experiment, and it is correlated with d (as mentioned earlier 60% of individuals undertook training when they were offered). The IV strategy is based on Chernozhukov and Hansen (2006, 2008) by considering the model

$$y - d\alpha = x\beta + z\gamma + u.$$

The IV method in QR proceeds as follows. Note that z does not belong to the model, as conditional on d , undertaking training, the offer has no effect on wages. Then, we construct a grid in $\alpha \in \mathcal{A}$, which is indexed by j for each $\tau \in (0, 1)$ and we estimate the quantile regression model for fixed τ

$$y - d\alpha_j(\tau) = x\beta + z\gamma + u.$$

This gives $\{\hat{\beta}_j(\alpha_j(\tau), \tau), \hat{\gamma}_j(\alpha_j(\tau), \tau)\}$, the set of conditional quantile regression estimates for the new model. Next, we choose α by minimizing a given norm of γ (we use the Euclidean norm),

$$\hat{\alpha}(\tau) = \underset{\alpha \in \mathcal{A}}{\operatorname{argmin}} \|\hat{\gamma}(\alpha(\tau), \tau)\|.$$

Figure 5 shows the values of γ^2 for the grids of α and τ . As a result we obtain the map $\tau \mapsto \{\hat{\alpha}(\tau), \hat{\beta}(\hat{\alpha}(\tau), \tau) \equiv \hat{\beta}(\tau), \hat{\gamma}(\hat{\alpha}(\tau), \tau) \equiv \hat{\gamma}(\tau)\}$.

[Figures 5]

Finally, we select the most probably quantile as in the previous case, by using the first order

condition corresponding the selection of τ :

$$\hat{\tau} = \underset{\tau \in (0,1)}{\operatorname{argmin}} \left| \frac{1 - 2\tau}{\tau(1 - \tau)} - \frac{\sum_{i=1}^n \hat{u}_i(\tau)}{\sum_{i=1}^n \rho_{\tau}(\hat{u}_i(\tau))} \right|$$

where $\hat{u}_i(\tau) = y_i - d_i \hat{\alpha}(\tau) - x'_i \hat{\beta}(\tau) - z_i \hat{\gamma}(\tau)$. Figure 6 reports the IV estimates together with the most likely quantile. Interestingly, the qualitative results are very much alike those of the value of the JTPA training offer. The IV least squares estimator for the effect of JTPA training gives a value of 0.116 (0.045) while IV median regression gives a much higher value of 0.142 (0.047). The most likely quantile continues to be 0.84 (0.053), which has an associated training effect of 0.072 (0.033). The ZQR effect continues to be smaller than the mean and median estimates. Therefore, the upper quantiles are more informative when analyzing the effects of JTPA training on log wages.

[Figure 6]

6 Conclusions

In this paper we show that the maximum likelihood problem for the asymmetric Laplace distribution can be found as the solution of a maximum entropy problem where we impose moment constraints given by the joint consideration of the mean and the median. We also propose an approximate Z-estimator method, which provides a parsimonious estimator that complements the quantile process. This provides an alternative interpretation of quantile regression and frames it within the maximum entropy paradigm. Potential estimates from this method has important applications. As an illustration, we apply the proposed estimator to a well-known dataset where quantile regression has been extensively used.

Appendix

A. Interpretation of the Z-estimator

In order to interpret θ_0 , we take the expectation of the estimating equations with respect to the unknown true density. To simplify the exposition we consider a simple model without covariates: $y_i = \alpha + u_i$. Our estimating equation vector is defined as:

$$E(\Psi_\theta(y)) = E \left(\begin{pmatrix} \frac{1}{\sigma} (\tau - 1(y < \alpha)) \\ \left(\frac{1-2\tau}{\tau(1-\tau)} - \frac{(y-\alpha)}{\sigma} \right) \\ -\frac{1}{\sigma} + \frac{1}{\sigma^2} \rho_\tau(y - \alpha) \end{pmatrix} \right) = 0,$$

and the estimator is such that

$$\frac{1}{n} \sum_{i=1}^n \Psi_\theta(y_i) = 0$$

Let $F(y)$ be the cdf of the random variable y . Now we need to find $E[\Psi_\theta(y)]$.

For the first component we have

$$\begin{aligned} \frac{1}{\sigma} E[\tau - I(y < \alpha)] &= \frac{1}{\sigma} \left(\int_{\mathbb{R}} (\tau - 1(y < \alpha)) dF(y) \right) \\ &= \frac{1}{\sigma} \left(\tau - \int_{-\infty}^{\alpha} dF(y) \right) \\ &= \frac{1}{\sigma} (\tau - F(\alpha)). \end{aligned}$$

Thus if we set this equal to zero, we have

$$\alpha = F^{-1}(\tau),$$

which is the usual quantile. Thus, the interpretation of the parameter α is analogous to QR if covariates are included.

For the third term in the vector, $-\frac{1}{\sigma} + \frac{1}{\sigma^2} \rho_\tau(y - \alpha)$, we have

$$E \left[-\frac{1}{\sigma} + \frac{1}{\sigma^2} \rho_\tau(y - \alpha) \right] = 0,$$

that is,

$$\sigma = E [\rho_\tau(y - \alpha)].$$

Thus, as in the least squares case, the scale parameter σ can be interpreted as the expected

value of the loss function.

Finally, we can interpret τ using the second equation,

$$E \left[\frac{1 - 2\tau}{\tau(1 - \tau)} - \frac{(y - \alpha)}{\sigma} \right] = 0,$$

which implies that

$$\frac{1 - 2\tau}{\tau(1 - \tau)} = \frac{E[y] - F^{-1}(\tau)}{\sigma}.$$

Note that $g(\tau) \equiv \frac{1-2\tau}{\tau(1-\tau)}$ is a measure of the skewness of the distribution (see also footnote 2). Thus, τ should be chosen to set $g(\tau)$ equal to a measure of asymmetry of the underline distribution $F(\cdot)$ given by the difference of τ -quantile with the mean (and standardized by σ). In the special case of a symmetric distribution, the mean coincides with the median and mode, such that $E[y] = F^{-1}(1/2)$ and $\tau = 1/2$, which is the most probable quantile and a solution to our Z-estimator.

B. Lemma A1

In this appendix we state an auxiliary result that states Donskerness and stochastic equicontinuity. Let $\mathcal{F} \equiv \{\psi_\theta(y, x), \theta \in \Theta\}$, and define the following empirical process notation for $w = (y, x)$:

$$f \mapsto \mathbb{E}_n[f(w)] = \frac{1}{n} \sum_{i=1}^n f(w_i) \quad f \mapsto \mathbb{G}_n[f(w)] = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(w_i) - Ef(w_i)).$$

We follow the literature using empirical process exploiting the monotonicity and boundedness of the indicator function, the boundedness of the moments of x and y , and that the problem is a parametric one.

Lemma A1. Under Assumptions A1-A4 \mathcal{F} is Donsker. Furthermore,

$$\theta \mapsto \mathbb{G}_n \psi_\theta(y, x)$$

is stochastically equicontinuous, that is

$$\sup_{\|\theta - \theta_0\| \leq \delta_n} \left\| \mathbb{G}_n \psi_\theta(y, x) - \mathbb{G}_n \psi_{\theta_0}(y, x) \right\| = o_p(1),$$

for any $\delta_n \downarrow 0$.

Proof: Note that a class \mathcal{F} of a vector-valued functions $f : x \mapsto \mathbb{R}^k$ is Donsker if each of the classes of coordinates $f_i : x \mapsto \mathbb{R}^k$ with $f = (f_1, \dots, f_k)$ ranging over \mathcal{F} ($i = 1, 2, \dots, k$) is Donsker (van der Vaart, 1998, p.270).

The first element of the vector is $\psi_{1\theta}(y, x) = (\tau - 1(y_i < x'_i\beta))\frac{x_i}{\sigma}$. Note that the functional class $\mathfrak{A} = \{\tau - 1\{y_i < x'_i\beta\}, \tau \in \mathcal{T}, \beta \in \mathcal{B}\}$ is a VC subgraph class and hence also Donsker class, with envelope 2. Its product with x also forms a Donsker class with a square integrable envelope $2 \cdot \max_j |x_j|$, by Theorem 2.10.6 in van der Vaart and Wellner (1996) (VW henceforth). Finally, the class \mathcal{F}_1 is defined as the product of the latter with $1/\sigma$, which is bounded. Thus, by assumption A4 \mathcal{F}_1 is Donsker. Now define the process $h_1 = (\beta, \tau, \sigma) \mapsto \mathbb{G}_n \psi_{1\theta}(y, x)$. Using the established Donskerness property, this process is Donsker in $l^\infty(\mathcal{F}_1)$.

The second element of the vector is $\psi_{2\theta}(y, x) = \left(\frac{1-2\tau}{\tau(1-\tau)} - \frac{(y_i - x'_i\beta)}{\sigma}\right)$. Define $\mathfrak{H} = \{(y_i - x'_i\beta), \beta \in \mathcal{B}\}$. Note that

$$|(y_i - x'_i\beta_1) - (y_i - x'_i\beta_2)| = |x'_i(\beta_2 - \beta_1)| \leq \|x_i\| \|\beta_2 - \beta_1\|,$$

where the inequality follows from Cauchy-Schwartz inequality. Thus by Assumptions A3-A4 and Example 19.7 in van der Vaart (1998) the class \mathfrak{H} is Donsker. Moreover, \mathfrak{H} belongs to a VC class satisfying a uniform entropy condition, since this class is a subset of the vector space of functions spanned by (y, x_1, \dots, x_p) , where p is the fixed dimension of x , so Lemma 2.6.15 of VW shows the desired result. Thus, by Example 2.10.23 (and Theorem 2.10.20) in VW the class defined by $1/\sigma \mathfrak{H}$ is Donsker, because the envelope of \mathfrak{H} ($|y| + \text{const} * |x|$) is square integrable by assumptions A3-A4. Thus \mathcal{F}_2 is Donsker. Using the same arguments as in the previous case we can define $h_2 = (\beta, \tau, \sigma) \mapsto \mathbb{G}_n \psi_{2\theta}(y, x)$, and by the established Donskerness property, this process is Donsker in $l^\infty(\mathcal{F}_2)$.

The third element of the vector is $\psi_{3\theta}(y, x) = \left(-\frac{1}{\sigma} + \frac{1}{\sigma^2} \rho_\tau(y_i - x'_i\beta)\right)$. Consider the following empirical process defined by $\mathfrak{J} = \{\rho_\tau(y_i - x'_i\beta), \tau \in \mathcal{T}, \beta \in \mathcal{B}\}$. This is Donsker by an application of Theorem 2.10.6 in VW. Finally, as in the previous cases define $h_3 = (\beta, \tau, \sigma) \mapsto \mathbb{G}_n \psi_{3\theta}(y, x)$, and by the established Donskerness property, this process is Donsker in $l^\infty(\mathcal{F}_3)$.

Now we turn our attention to the stochastic equicontinuity. The process $\theta \mapsto \mathbb{G}_n \psi_\theta(y, x)$ is stochastically equicontinuous over Θ with respect to a $L_2(P)$ pseudometric.⁶ First, as in

⁶See e.g. Kosorok (2008, p. 405) for a sufficient condition for stochastic equicontinuity.

Angrist, Chernozhukov, and Fernández-Val (2006) and Chernozhukov and Hansen (2006), we define the distance d as the following $L_2(P)$ pseudometric

$$d(\theta', \theta'') = \sqrt{E([\psi_{\theta'} - \psi_{\theta''}]^2)}.$$

Thus, as $\|\theta - \theta_0\| \rightarrow 0$ we need to show that

$$d(\theta, \theta_0) \rightarrow 0, \tag{28}$$

and therefore, by Donskerness of $\theta \mapsto \mathbb{G}_n \Psi_\theta(y, x)$, we have

$$\mathbb{G}_n \psi_\theta(y, x) = \mathbb{G}_n \psi_{\theta_0}(y, x) + o_p(1),$$

that is

$$\sup_{|\theta - \theta_0| \leq \delta_n} \|\mathbb{G}_n \psi_\theta(y, x) - \mathbb{G}_n \psi_{\theta_0}(y, x)\| = o_p(1).$$

To show (28), first note that

$$\begin{aligned} d(\theta', \theta) &= \sqrt{E([\psi_{1\theta'} - \psi_{1\theta}]^2)} \\ &= \sqrt{E\left(\left[\left(\tau' - 1(y - x\beta')\right)\frac{x}{\sigma'} - \left(\tau - 1(y - x\beta)\right)\frac{x}{\sigma}\right]^2\right)} \\ &\leq \left[\left(E \left| \frac{1}{\sigma'}(\tau' - 1(y - x\beta')) - \frac{1}{\sigma}1(\tau - (y - x\beta)) \right|^{\frac{2(2+\epsilon)}{\epsilon}} \right)^{\frac{\epsilon}{2(2+\epsilon)}} \cdot \left(E(|x|^2)^{\frac{2+\epsilon}{2}} \right)^{\frac{2}{2+\epsilon}} \right]^{\frac{1}{2}} \\ &= \left(E \left| \left(\frac{\tau'}{\sigma'} - \frac{\tau}{\sigma} \right) + \left(\frac{1}{\sigma}1(y \leq x\beta) - \frac{1}{\sigma'}1(y \leq x\beta') \right) \right|^{\frac{2(2+\epsilon)}{\epsilon}} \right)^{\frac{\epsilon}{2(2+\epsilon)}} \cdot \left(E(|x|^2)^{\frac{2+\epsilon}{2}} \right)^{\frac{1}{2+\epsilon}} \\ &\leq \left[\left(E \left(\left| \frac{\tau'}{\sigma'} - \frac{\tau}{\sigma} \right| \right)^{\frac{2(2+\epsilon)}{\epsilon}} \right)^{\frac{\epsilon}{2(2+\epsilon)}} + \left(E \left(\left| \frac{1}{\sigma}1(y \leq x\beta) - \frac{1}{\sigma'}1(y \leq x\beta') \right| \right)^{\frac{2(2+\epsilon)}{\epsilon}} \right)^{\frac{\epsilon}{2(2+\epsilon)}} \right] \cdot \left(E(|x|^2)^{\frac{2+\epsilon}{2}} \right)^{\frac{1}{2+\epsilon}} \\ &\leq \left[\left| \frac{\tau'}{\sigma'} - \frac{\tau}{\sigma} \right| + \left(E \left| \bar{g} \cdot x' \left(\frac{\beta'}{\sigma'} - \frac{\beta}{\sigma} \right) \right|^{\frac{\epsilon}{2(2+\epsilon)}} \right) \right] \cdot \left(E\|x\|^{2+\epsilon} \right)^{\frac{1}{2+\epsilon}} \\ &\leq \left[\left| \frac{\tau'}{\sigma'} - \frac{\tau}{\sigma} \right| + \left(\bar{g} E\|x\| \left\| \frac{\beta'}{\sigma'} - \frac{\beta}{\sigma} \right\| \right)^{\frac{\epsilon}{2(2+\epsilon)}} \right] \cdot \left(E\|x\|^{2+\epsilon} \right)^{\frac{1}{2+\epsilon}}, \end{aligned}$$

where the first inequality is Holder's inequality, the second is Minkowski's inequality, the

third is a Taylor expansion as in Angrist, Chernozhukov, and Fernández-Val (2006) where \bar{g} is the upper bound of $g(y|x)$ (using A2), and the last is Cauchy-Schwarz inequality.

Now rewrite $\psi_{2\theta}(y, x) = \left(\sigma \frac{1-2\tau}{\tau(1-\tau)} - (y - x'\beta) \right)$ and

$$\begin{aligned}
d(\theta', \theta) &= \sqrt{E([\psi_{2\theta'} - \psi_{2\theta}]^2)} \\
&= \sqrt{E\left(\left[\sigma' \frac{1-2\tau'}{\tau'(1-\tau')} - (y - x'\beta') - \sigma \frac{1-2\tau}{\tau(1-\tau)} + (y - x'\beta)\right]^2\right)} \\
&= \sqrt{E\left(\left|\sigma' \frac{1-2\tau'}{\tau'(1-\tau')} - \sigma \frac{1-2\tau}{\tau(1-\tau)} + x'(\beta - \beta')\right|^2\right)} \\
&\leq \left(E\left|\sigma' \frac{1-2\tau'}{\tau'(1-\tau')} - \sigma \frac{1-2\tau}{\tau(1-\tau)}\right|^2\right)^{1/2} + (E|x'(\beta - \beta')|^2)^{1/2} \\
&\leq \left(E\left|\sigma' \frac{1-2\tau'}{\tau'(1-\tau')} - \sigma \frac{1-2\tau}{\tau(1-\tau)}\right|^2\right)^{1/2} + \|\beta' - \beta\| (E\|x\|^2)^{1/2},
\end{aligned}$$

where the first inequality is given by Minkowski's inequality $(E|X + Y|^p)^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$ for $p \geq 1$, and the second inequality is Cauchy-Schwarz inequality.

Finally, rewrite $\psi_{3\theta}(y, x) = (-\sigma + \rho_\tau(y - x'\beta))$, and thus

$$\begin{aligned}
d(\theta', \theta) &= \sqrt{E([\psi_{3\theta'} - \psi_{3\theta}]^2)} \\
&= \sqrt{E\left(\left[-\sigma' + \rho_{\tau'}(y - x\beta') + \sigma - \rho_\tau(y - x\beta)\right]^2\right)} \\
&= \sqrt{E\left(\left[-\sigma' + \sigma + \rho_{\tau'}(y - x\beta') - \frac{1}{\sigma^2}\rho_\tau(y - x\beta)\right]^2\right)} \\
&\leq \sqrt{E(-\sigma' + \sigma)^2} + \sqrt{E\left(\left[\rho_{\tau'}(y - x\beta') - \rho_\tau(y - x\beta)\right]^2\right)} \\
&= \sqrt{E(-\sigma' + \sigma)^2} + \sqrt{E\left(\left[\rho_{\tau'}(y - x\beta') - \rho_{\tau'}(y - x\beta) + \rho_{\tau'}(y - x\beta) - \rho_\tau(y - x\beta)\right]^2\right)} \\
&\leq |\sigma - \sigma'| + \sqrt{E\left(\left[\|x(\beta' - \beta)\| + |\tau' - \tau|(y - x\beta)\right]^2\right)} \\
&\leq |\sigma - \sigma'| + \sqrt{E\left(\left[\|x\|\|\beta' - \beta\| + |\tau' - \tau|(y - x\beta)\right]^2\right)} \\
&\leq |\sigma - \sigma'| + \left(E[\|x\|\|\beta' - \beta\|^2]\right)^{1/2} + \left(E[|\tau' - \tau|(y - x\beta)]^2\right)^{1/2} \\
&= |\sigma - \sigma'| + \|\beta' - \beta\| (E\|x\|^2)^{1/2} + |\tau' - \tau| (E[(y - x\beta)]^2)^{1/2} \\
&\leq \text{const} \cdot (|\sigma - \sigma'| + \|\beta' - \beta\| + |\tau' - \tau|),
\end{aligned}$$

where the first inequality is given by Minkowski's inequality, the second inequality is given by QR check function properties as $\rho_\tau(x + y) - \rho_\tau(y) \leq 2|x|$ and $\rho_{\tau_1}(y - x't) - \rho_{\tau_2}(y - x't) = (\tau_2 - \tau_1)(y - x't)$. Third inequality is Cauchy-Schwarz inequality. Fourth is Minkowski's inequality. Last inequality uses assumption A4.

Thus, $\|\theta' - \theta\| \rightarrow 0$ implies that $d(\theta', \theta) \rightarrow 0$ in every case, and therefore, by Donskerness of $\theta \mapsto \mathbb{G}_n \psi_\theta(y, x)$ we have that

$$\sup_{\|\theta - \theta_0\| \leq \delta_n} \|\mathbb{G}_n \psi_\theta(y, x) - \mathbb{G}_n \psi_{\theta_0}(y, x)\| = o_p(1).$$

■

References

- ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings,” *Econometrica*, 70, 91–117.
- ANGRIST, J., V. CHERNOZHUKOV, AND I. FERNÁNDEZ-VAL (2006): “Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure,” *Econometrica*, 74, 539–563.
- BLOOM, H. S. B., L. L. ORR, S. H. BELL, G. CAVE, F. DOOLITTLE, W. LIN, AND J. M. BOS (1997): “The Benefits and Costs of JTPA Title II-A programs. Key Findings from the National Job Training Partnership Act Study,” *Journal of Human Resources*, 32, 549–576.
- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND B. MELLY (2009): “Inference on counterfactual distributions,” CEMMAP working paper CWP09/09.
- CHERNOZHUKOV, V., AND C. HANSEN (2006): “Instrumental Quantile Regression Inference for Structural and Treatment Effects Models,” *Journal of Econometrics*, 132, 491–525.
- (2008): “Instrumental Variable Quantile Regression: A Robust Inference Approach,” *Journal of Econometrics*, 142, 379–398.
- EBRAHIMI, N., E. S. SOOFI, AND R. SOYER (2008): “Multivariate Maximum Entropy Identification, Transformation, and Dependence,” *Journal of Multivariate Analysis*, 99, 1217–1231.
- FIRPO, S. (2007): “Efficient Semiparametric Estimation of Quantile Treatment Effects,” *Econometrica*, 75, 259–276.
- GERACI, M., AND M. BOTAI (2007): “Quantile regression for longitudinal data using the asymmetric Laplace distribution,” *Biostatistics*, 8, 140–154.
- HE, X., AND Q.-M. SHAO (1996): “A General Bahadur Representation of M-Estimators and its Applications to Linear Regressions with Nonstochastic Designs,” *Annals of Statistics*, 24, 2608–2630.
- (2000): “Quantile regression estimates for a class of linear and partially linear errors-in-variables models,” *Statistica Sinica*, 10, 129–140.

- HINKLEY, D. V., AND N. S. REVANKAR (1977): “Estimation of the Pareto Law from Underreported Data: A Further Analysis,” *Journal of Econometrics*, 5, 1–11.
- HUBER, P. J. (1967): “The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions,” in *Fifth Symposium on Mathematical Statistics and Probability*, pp. 179–195. University of California, Berkeley, California.
- KITAMURA, Y., AND M. STUTZER (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation,” *Econometrica*, 65, 861–874.
- KOENKER, R. (2005): *Quantile Regression*. Cambridge University Press, Cambridge.
- KOENKER, R., AND G. W. BASSETT (1978): “Regression Quantiles,” *Econometrica*, 46, 33–49.
- KOENKER, R., AND J. A. F. MACHADO (1999): “Godness of Fit and Related Inference Processes for Quantile Regression,” *Journal of the American Statistical Association*, 94, 1296–1310.
- KOENKER, R., AND Z. XIAO (2002): “Inference on the Quantile Regression Process,” *Econometrica*, 70, 1583–1612.
- KOMUNJER, I. (2005): “Quasi-maximum likelihood estimation for conditional quantiles,” *Journal of Econometrics*, 128, 137–164.
- (2007): “Asymmetric Power Distribution: Theory and Applications to Risk Measurement,” *Journal of Applied Econometrics*, 22, 891–921.
- KOSOROK, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag Press, New York, New York.
- KOTZ, S., T. J. KOZUBOWSKI, AND K. PODGÓRSKI (2002a): “Maximum Entropy Characterization of Asymmetric Laplace Distribution,” *International Mathematical Journal*, 1, 31–35.
- (2002b): “Maximum Likelihood Estimation of Asymmetric Laplace Distributions,” *Annals of the Institute Statistical Mathematics*, 54, 816–826.
- LALONDE, R. J. (1995): “The Promise of Public-Sponsored Training Programs,” *Journal of Economic Perspectives*, 9, 149–168.
- MACHADO, J. A. F. (1993): “Robust Model Selection and M-Estimation,” *Econometric Theory*, 9, 478–493.

- MANSKI, C. F. (1991): “Regression,” *Journal of Economic Literature*, 29, 34–50.
- PARK, S. Y., AND A. K. BERA (2009): “Maximum entropy autoregressive conditional heteroskedasticity model,” *Journal of Econometrics*, 150, 219–230.
- SCHENNACH, S. M. (2008): “Quantile Regression with Mismeasured Covariates,” *Econometric Theory*, 24, 1010–1043.
- SOOFI, E. S., AND J. J. RETZER (2002): “Information Indices: Unification and Applications,” *Journal of Econometrics*, 107, 17–40.
- VAN DER VAART, A. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- VAN DER VAART, A., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag, New York, New York.
- WEI, Y., AND R. J. CARROLL (2009): “Quantile regression with measurement error,” *Journal of the American Statistical Association*, 104, 1129–1143.
- YU, K., AND R. A. MOYEED (2001): “Bayesian quantile regression,” *Statistics & Probability Letters*, 54, 437–447.
- YU, K., AND J. ZHANG (2005): “A three-parameter asymmetric Laplace distribution and its extension,” *Communications in Statistics - Theory and Methods*, 34, 1867–1879.

Table 1: Location-Shift Model: Bias and RMSE

		ZQR	QR (Median)	OLS
$N(0, 1)$	Bias	0.0007	-0.0004	0.0008
	RMSE	0.0904	0.0899	0.0710
	$\hat{\tau}$	0.501	-	-
t_3	Bias	0.0012	-0.0008	0.0014
	RMSE	0.1133	0.0967	0.1217
	$\hat{\tau}$	0.498	-	-
χ_3^2	Bias	-0.0021	0.0024	0.0020
	RMSE	0.1419	0.1892	0.1801
	$\hat{\tau}$	0.081	-	-
ALPD ($\tau = 0.5$)	Bias	0.0001	0.0001	0.0001
	RMSE	0.0638	0.0549	0.0710
	$\hat{\tau}$	0.499	-	-
ALPD ($\tau = 0.25$)	Bias	-0.0008	-0.0001	0.0003
	RMSE	0.0718	0.0860	0.0917
	$\hat{\tau}$	0.248	-	-

Table 2: Location-Scale-Shift Model: Bias and RMSE

		ZQR	QR (Median)	OLS
$N(0, 1)$	Bias	0.0015	0.0036	0.0037
	RMSE	0.2209	0.1461	0.1365
	$\hat{\tau}$	0.499	-	-
t_3	Bias	-0.0005	0.0002	-0.0052
	RMSE	0.2457	0.1460	0.2565
	$\hat{\tau}$	0.501	-	-
χ_3^2	Bias	-0.0004	0.0076	0.0089
	RMSE	0.5087	0.2833	0.3788
	$\hat{\tau}$	0.086	-	-
ALPD ($\tau = 0.5$)	Bias	-0.0010	-0.0001	-0.0013
	RMSE	0.1455	0.0845	0.1459
	$\hat{\tau}$	0.501	-	-
ALPD ($\tau = 0.25$)	Bias	0.0051	0.0004	0.4076
	RMSE	0.1331	0.1429	0.4505
	$\hat{\tau}$	0.248	-	-

Table 3: JTPA offer

	ZQR [$\hat{\tau} = 0.84$]		OLS		Median regression	
Intercept	9.894	(0.059)	8.814	(0.088)	9.188	(0.086)
JTPA offer	0.045	(0.022)	0.075	(0.032)	0.100	(0.033)
FEMALE	0.301	(0.023)	0.259	(0.030)	0.260	(0.031)
HSORGED	0.201	(0.025)	0.267	(0.034)	0.297	(0.037)
BLACK	-0.102	(0.026)	-0.121	(0.036)	-0.175	(0.039)
HISPANIC	-0.032	(0.034)	-0.034	(0.050)	-0.025	(0.051)
MARRIED	0.129	(0.025)	0.242	(0.036)	0.265	(0.034)
WKLESS13	-0.255	(0.023)	-0.598	(0.032)	-0.556	(0.036)
AGE2225	0.229	(0.057)	0.175	(0.084)	0.125	(0.080)
AGE2629	0.285	(0.058)	0.192	(0.085)	0.131	(0.081)
AGE3035	0.298	(0.057)	0.191	(0.084)	0.176	(0.080)
AGE3644	0.320	(0.058)	0.130	(0.085)	0.173	(0.081)
AGE4554	0.267	(0.064)	0.110	(0.094)	0.080	(0.092)
$\hat{\tau}$	0.840	(0.051)			0.500	
$\hat{\sigma}$	0.249	(0.060)			0.538	(0.006)

Notes: 9872 observations. The numbers in parenthesis are the corresponding standard errors. JTPA offer: dummy variable for individuals that received a JTPA offer; FEMALE: Female dummy variable; HSORGED: dummy variable for individuals with completed high school or GSE; BLACK: race dummy variable; HISPANIC: dummy variable for hispanic; MARRIED: dummy variable for married individuals; WKLESS13: dummy variable for individuals working less than 13 weeks in the past year; AGE2225, AGE2629, AGE3035, AGE3644 and AGE4554 age range indicator variables.

Figure 1: Linear Combination of $|y - \mu|$ and $(y - \mu)$

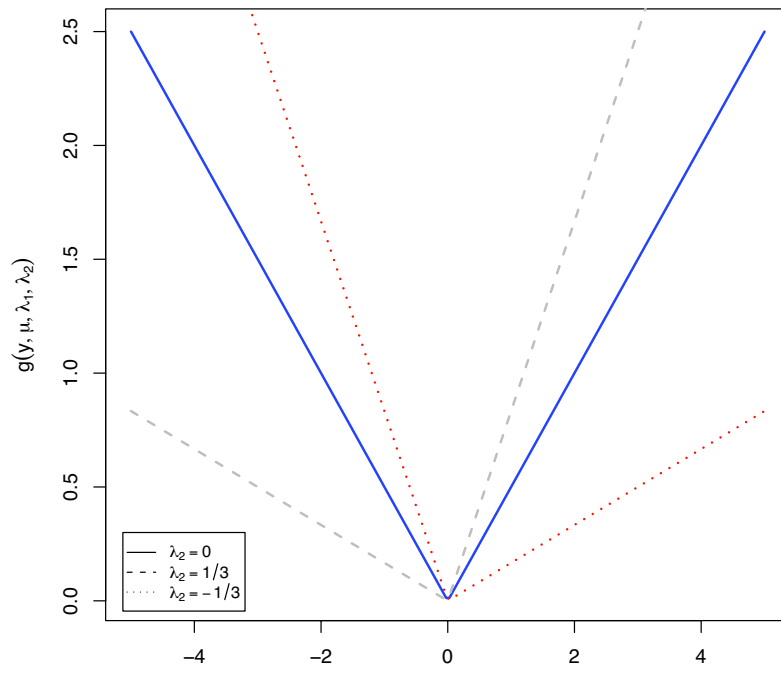
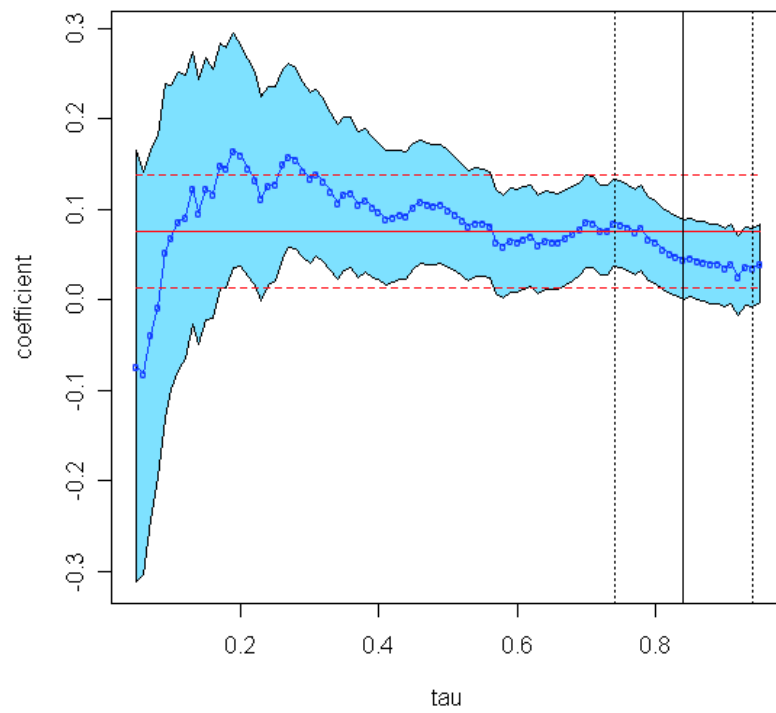
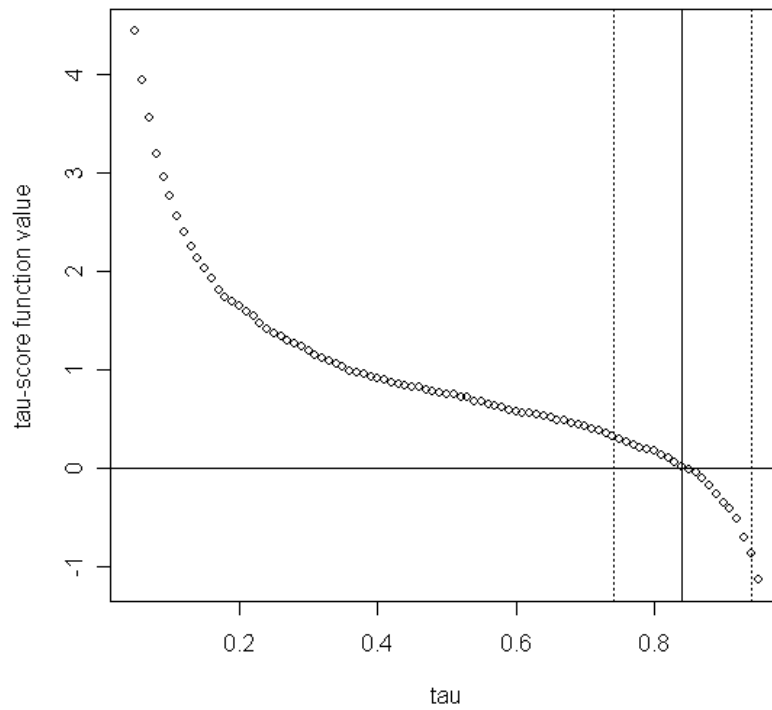


Figure 2: JTPA offer: Quantile regression process and OLS



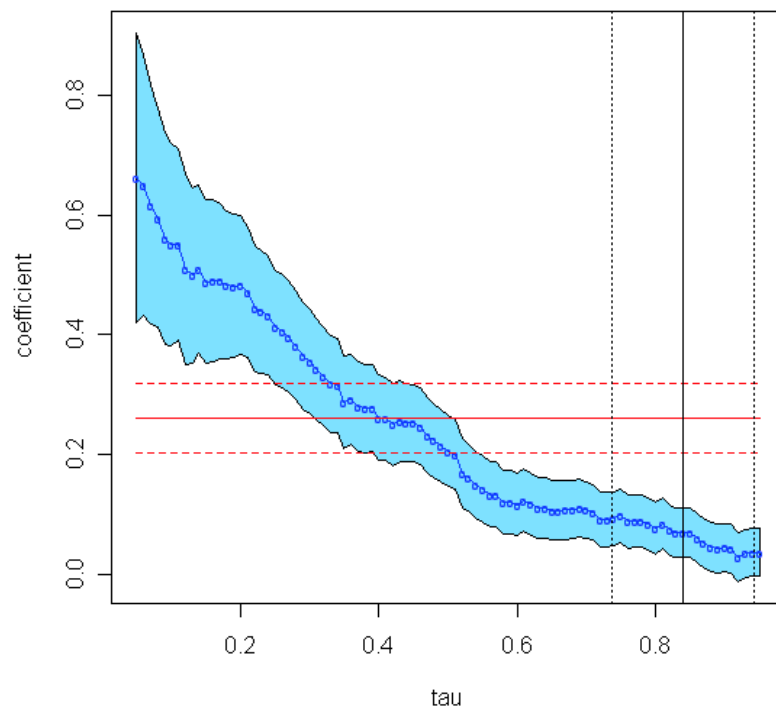
Notes: Quantile regression process (shaded area), OLS (horizontal lines) and estimated most informative quantile (vertical lines) with 95% confidence intervals.

Figure 3: JTPA offer: τ -score function



Notes: The τ -score function is $\frac{1-2\tau}{\tau(1-\tau)} - \frac{\sum_{i=1}^n (y_i - x_i' \hat{\beta}(\tau))}{n\hat{\sigma}}$.

Figure 4: JTPA: Quantile regression process and OLS



Notes: Quantile regression process (shaded area), OLS (horizontal lines) and estimated most informative quantile (vertical lines) with 95% confidence intervals.

Figure 5: JTPA: Minimization of $\|\gamma^2(\tau, \alpha)\|$

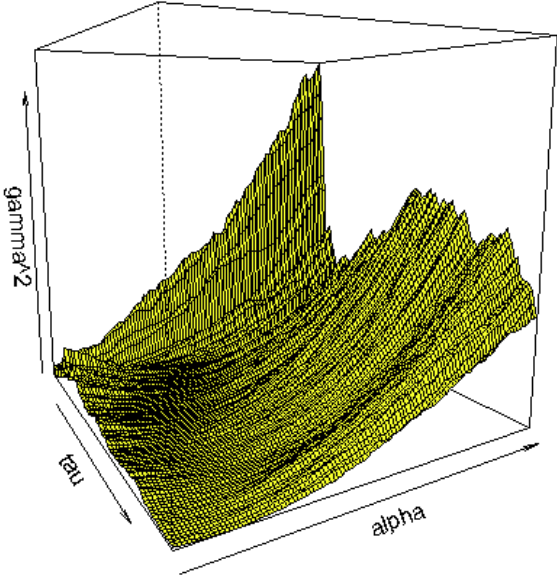
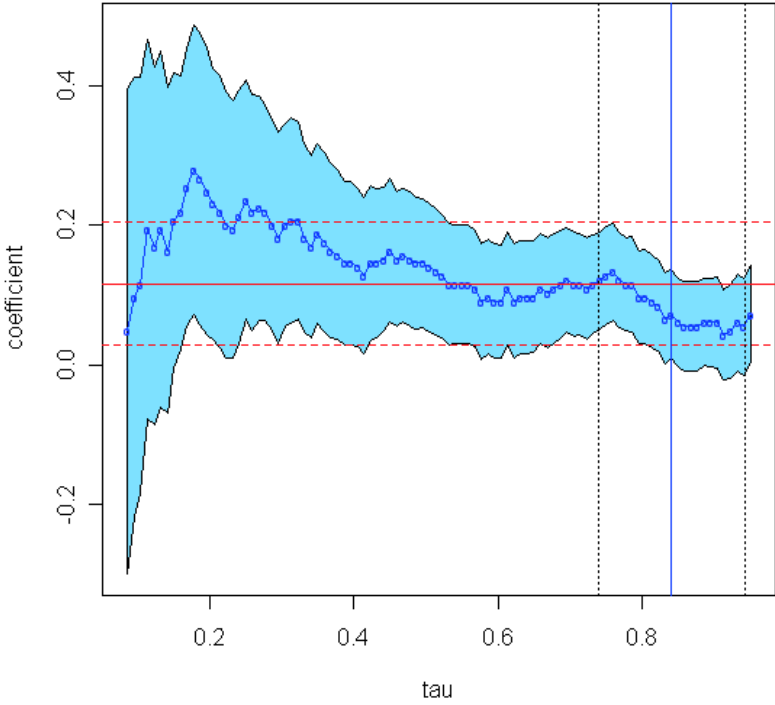


Figure 6: JTPA: IV Quantile regression process and IV OLS



Notes: Quantile regression process (shaded area), OLS (horizontal lines) and estimated most informative quantile (vertical lines) with 95% confidence intervals.