Instrumental Variable Quantile Regression[*] [†]

Victor Chernozhukov and Christian Hansen

First Version: May 2001 Revised: December 2004

Abstract

Quantile regression is an increasingly important tool that estimates the conditional quantiles of a response Y given a vector of regressors D. It usefully generalizes Laplace's median regression and can be used to measure the effect of covariates not only in the center of a distribution, but also in the upper and lower tails. For the linear quantile model defined by $Y = D'\gamma(U)$ where $D'\gamma(U)$ is strictly increasing in U and U is a standard uniform variable independent of D, quantile regression allows estimation of quantile specific covariate effects $\gamma(\tau)$ for $\tau \in (0,1)$. In this paper, we propose an instrumental variable quantile regression estimator that appropriately modifies the conventional quantile regression and recovers quantile-specific covariate effects in an instrumental variables model defined by $Y = D'\alpha(U)$ where $D'\alpha(U)$ is strictly increasing in U and U is a uniform variable that may depend on D but is independent of a set of instrumental variables Z. The proposed estimator and inferential procedures are computationally convenient in typical applications and can be carried out using software available for conventional quantile regression. In addition, the proposed estimation procedure gives rise to a convenient inferential procedure that is naturally robust to weak identification. The use of the proposed estimator and testing procedure is illustrated through two empirical examples.

Keywords: Quantile Regression, Instrumental Variables, Weak Instruments

1. Introduction

Quantile regression is an important tool for estimating conditional quantile models that has been used in many empirical studies and has been studied extensively in theoretical statistics; see Koenker and Bassett (1978), Koenker and Portnoy (1987), Portnoy (1991), Gutenbrunner and Jurečková (1992), Chaudhuri, Doksum, and Samarov (1997), Portnoy and Koenker (1997), Knight (1998), Koenker and Machado (1999), Portnoy (2001), and He and Zhu (2003). One of quantile regression's

[*]The Matlab software for this paper is available upon request via e-mail chansen1@gsb.uchicago.edu. Further updates of this paper can be downloaded at www.mit.edu/˜vchern. Address correspondence to C. Hansen, Asst. Prof. of Econometrics and Statistics, The University of Chicago, Graduate School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA, chansen1@gsb.uchicago.edu.

[†]Portions of this paper were previously included in MIT Department of Economics Working Paper 02-07, "An IV Model of Quantile Treatment Effects", 2001.

most appealing features is its ability to estimate quantile-specific effects that describe the impact of covariates not only on the center but also on the tails of the outcome distribution. While the central effects, such as the mean effect obtained through conditional mean regression, provide interesting summary statistics of the impact of a covariate, they fail to describe the full distributional impact unless the variable affects both the central and the tail quantiles in the same way. In addition, interest focuses on the impact of covariates on points other than the center of the distribution in many cases. For example, in a study of the effectiveness of a job training program, the effect of training on the low tail of the earnings distribution will likely be of more interest for public policy than the effect of training on the mean of the distribution.

For an outcome Y and set of factors D affecting the outcome, the conventional linear conditional quantile model may be defined as

$$Y = D'\gamma(U^*), \qquad U^*|D \sim \text{Uniform}(0,1), \tag{1.1}$$

where $\tau \mapsto D'\gamma(\tau)$ is strictly increasing and continuous in τ. Doksum (1974) interprets the disturbance U^* as individual ability or proneness. By construction, $D'\gamma(\tau)$ is the τ-quantile of Y conditional on D. This model generalizes the usual linear regression model

$$Y = D'\gamma_0 + \gamma_1(U^*)$$

by allowing quantile-specific effects of covariates D. For a given quantile indexed by $\tau \in (0,1)$, the quantile specific effects $\gamma(\tau)$ can be estimated using standard quantile regression methods (e.g. Koenker and Bassett (1978)).

In this paper, we develop a new estimation method for an endogenous generalization of the above model. The developed approach is designed for settings where the observed variables D are determined non-experimentally, making it difficult to infer the true structural/causal effects of D on the outcomes. Specifically, we consider the model

$$Y = D'\alpha(U), \qquad U|Z \sim \text{Uniform}(0,1), \tag{1.2}$$

where $\tau \mapsto D'\alpha(\tau)$ is strictly increasing in τ, D is statistically dependent on U, and Z is a set of instrumental variables that are independent of U but statistically related to D. Since D depends on U, the sampled D will depend on U, leading to biased sampling or endogeneity. This endogeneity makes $\gamma(\tau) \neq \alpha(\tau)$, rendering conventional quantile regression inconsistent for estimating (1.2).

For example, suppose that Y is the hourly wage of a worker and that D is an individual's level of training. The unobserved disturbance U would reflect unobserved personal characteristics, such as ability, which influence the individual's wage via the equation $Y = D'\alpha(U)$. If high-ability individuals choose high levels of training, then the level of training is correlated with ability, which causes dependence between U and D and implies that conventional quantile regression will overstate the true effect of training on earnings, $\gamma(\tau) > \alpha(\tau)$. Instrumental variables Z, such as random assignment to training programs in the training context, allow us to overcome this problem by providing a source of variation in D that is independent of U. There are many other interesting examples where D is sampled depending on U, i.e. endogenously. The empirical section presents a supply-demand example and a training example.

Model (1.2) generalizes the conventional instrumental variables model with additive disturbances $Y = D'\alpha_0 + \alpha_1(U)$ where $U|Z \sim U(0,1)$ to cases where the impact of D varies across quantiles of the outcome distribution. A number of appealing approaches are readily available to estimate α_0 in the conventional instrumental variables model with additive disturbances, including the conventional two-stage least squares (2SLS) estimator and its robust analogs by Amemiya (1982) and Chen and Portnoy (1996).

The purpose of this paper is to provide practical estimation and inference methods for model (1.2). The estimator we propose is an appealing modification of the standard quantile regression that can be constructed from a series of conventional quantile regressions. Thus, the estimation approach is computationally convenient and simple to implement in many typical applications. It has already been used in empirical applications, e.g. Hausman and Sidak (2004), Januszewski (2004), and Chernozhukov and Hansen (2004a), and will be further illustrated with two empirical applications in this paper. In addition, the estimation procedure leads naturally to an inference procedure that will be valid even when one of the key conditions for identification of the model, that D is statistically dependent on Z, fails.

The remainder of this paper is organized as follows. In Section 2, we define the model in more detail and allow for other controls in the equations. Section 3 discusses estimation and testing procedures based on a set of moment equations introduced in Section 2. Section 4 illustrates the use of the derived estimator and testing procedure through brief empirical examples, and Section 5 concludes.

2. The Instrumental Quantile Regression Method

2.1. The Model

In this section, we more formally define the model we will estimate. Suppose we have a structural relationship defined by

$$Y = D'\alpha(U) + X'\beta(U), \qquad U|X,Z \sim \text{Uniform}(0,1), \tag{2.1}$$

$$D = \delta(X, Z, V), \text{ where } V \text{ is statistically dependent on } U, \tag{2.2}$$

$$\tau \mapsto D'\alpha(\tau) + X'\beta(\tau) \text{ strictly increasing in } \tau. \tag{2.3}$$

In these equations,

- Y is the scalar outcome variable of interest,
- U is a scalar random variable that aggregates all of the unobserved factors affecting the structural outcome equation,
- D is a vector of endogenous variables determined by (2.2), where
- V is a vector of unobserved disturbances determining D and correlated with U,
- Z is a vector of instrumental variables (control variables excluded from (2.1) that are independent of the disturbance U but impact variable D via (2.2)), and
- X is a vector of included control variables.

The observed variables consist of (Y, D, X, Z), and due to the dependence between V and U, D is also sampled depending on U.

We shall refer to the function

$$S_Y(\tau|d,x) = d'\alpha(\tau) + x'\beta(\tau) \tag{2.4}$$

as the Structural Quantile Function (SQF) in order to emphasize that it is in general a different object than the conditional quantile function $Q_Y(\tau|d,x)$. The structural quantile function $S_Y(\tau|d,x)$ describes the quantile function of the latent outcome variable $Y_d = d'\alpha(U) + X'\beta(U)$ obtained by fixing $D = d$ and sampling the disturbance $U \sim U(0,1)$ (all conditional on X). This notion of sampling corresponds to independent sampling of D and U, which is generally not feasible outside experimental settings. Instead the sampled variable D is determined via (2.2). Nevertheless, it is still possible to estimate or make inference on the structural quantile function $S_Y(\tau|d,x)$ through the use of instrumental variables Z which induce variation in D but are themselves independent of U.

2.2. The Principle

Under the conditions of (2.1) and (2.3), the problem of dependence between U and D is overcome through the presence of instrumental variables, Z, that affect the determination of D but are independent of U. In program evaluation studies with imperfect compliance, a simple example of an instrument is random assignment to the treatment group, which is done independently of the potential values of U. The presence of the instrumental variable leads to a set of moment equations that can be used to estimate the parameters of (2.1). From (2.1) and (2.3), the event $\{Y \le S_Y(\tau|D,X)\}$ is equivalent to the event $\{U \le \tau\}$. It then follows from (2.1) that

$$P[Y \le S_Y(\tau|D,X)|Z,X] = \tau. \tag{2.5}$$

Equation (2.5) provides a useful statistical restriction that can be used to estimate the structural parameters α and β. It is important to notice that the equation $P[Y \leq S_Y(\tau|D,X)|Z,X] = \tau$ differs from the conventional estimating equation

$$P[Y \leq Q_Y(\tau|D,X)|D,X] = \tau \tag{2.6}$$

used to estimate the conditional quantile function of Y given D and X.

Recall from Koenker and Bassett (1978) that the ordinary quantile regression (QR) is formulated as finding the best predictor of Y given W under the asymmetric least absolute deviation loss $\rho_\tau(u) := (\tau - 1(u < 0))u$. In other words, assuming integrability, the τ-th conditional quantile of Y given W solves the problem:

$$Q_Y(\tau|W) = \arg\min_{f \in \mathcal{F}} E\left[\rho_\tau(Y - f(W))\right] \tag{2.7}$$

where $\mathcal{F}$ is the class of measurable functions of W (restricted in applications to be a set of flexible parametric functions). Laplace's median regression function $Q_Y(.5|W)$ is a solution of this problem with $\tau = 1/2$ so that $\rho_\tau(u) = \frac{1}{2}|u|$. The function $Q_Y(\tau|D,X)$ solves the above problem with $W = (D,X)$ and can be estimated using the finite sample analog of the above equation.

The moment equation given in (2.5) is equivalent to the statement that 0 is the τ-th quantile of random variable $Y - S_Y(\tau|D,X)$ conditional on (Z,X):

$$0 = Q_{Y - S_Y(\tau|D,X)}(\tau|Z,X) \quad \text{a.s. for each } \tau. \tag{2.8}$$

Thus, we may pose the problem of finding $\alpha(\tau)$ and $\beta(\tau)$ solving equation (2.5) as the **instrumental variable** or **inverse quantile regression** (IVQR). This problem is to find an $S_Y(\tau|D,X)$ such that $\mathbf{0}$ is a solution to the quantile regression of $Y - S_Y(\tau|D,X)$ on (Z,X):

$$0 = \arg\min_{f \in \mathcal{F}} E\rho_\tau\left[(Y - S_Y(\tau|D,X) - f(Z,X))\right], \tag{2.9}$$

where $\mathcal{F}$ is the class of measurable functions of (X,Z) (which will be restricted in applications). The term 'inverse' emphasizes an evident inverse relation of this problem to the conventional quantile regression, (2.7).

3. The Instrumental Quantile Regression Estimator and Derived Dual Inference

3.1. Basic Description and Properties

Next we consider a finite-sample analog of the above procedure. Define the (weighted) conventional quantile regression objective function as

$$Q_n(\tau, \alpha, \beta, \gamma) := \frac{1}{n} \sum_{i=1}^{n} \rho_\tau (Y_i - D_i' \alpha - X_i' \beta - \mathcal{Z}_i' \gamma) V_i,$$

where D is a $\dim(\alpha)$-vector of endogenous variables, X is a $\dim(\beta)$-vector of exogenous explanatory variables, $\mathcal{Z}_i \equiv f(X_i, Z_i)$ is a $\dim(\gamma)$-vector of instrumental variables such that $\dim(\gamma) \geq \dim(\alpha)$, and $V_i := V(X_i, Z_i) > 0$ is a scalar weight. In practice, a simple procedure is to set $V_i = 1$ and let $\mathcal{Z}_i$ either be Z_i or the predicted value from a least squares projection of D_i on Z_i and X_i.

The **instrumental variable** or **inverse quantile regression estimator** (IVQR) is defined as follows. For a given value of the structural parameter, say α, let us run the ordinary QR to obtain

$$(\widehat{\beta}(\alpha, \tau), \widehat{\gamma}(\alpha, \tau)) := \arg\min_{\beta, \gamma} \; Q_n(\tau, \alpha, \beta, \gamma). \tag{3.1}$$

To find an estimate for $\alpha(\tau)$, we will look for a value α that makes the coefficient on the instrumental variable $\widehat{\gamma}(\alpha, \tau)$ as close to 0 as possible. Formally, let

$$\widehat{\alpha}(\tau) = \arg \inf_{\alpha \in \mathcal{A}} [W_n(\alpha)], \quad W_n(\alpha) := n[\widehat{\gamma}(\alpha, \tau)'] \widehat{A}(\alpha) [\widehat{\gamma}(\alpha, \tau)], \tag{3.2}$$

where $\widehat{A}(\alpha) = A(\alpha) + o_p(1)$ and $A(\alpha)$ is positive definite, uniformly in $\alpha \in \mathcal{A}$. It is convenient to set $A(\alpha)$ equal to the inverse of the asymptotic covariance matrix of $\sqrt{n}(\widehat{\gamma}(\alpha, \tau) - \gamma(\alpha, \tau))$ in which case $W_n(\alpha)$ is the Wald statistic for testing $\gamma(\alpha, \tau) = 0$, a fact that we will use below for inference about $\alpha(\tau)$ itself. The parameter estimates are then given by

$$\widehat{\theta}(\tau) := \left(\widehat{\alpha}(\tau), \widehat{\beta}(\tau)\right) := \left(\widehat{\alpha}(\tau), \widehat{\beta}(\widehat{\alpha}(\tau), \tau)\right). \tag{3.3}$$

The estimator (3.3) is a finite-sample instrumental variable quantile regression. Analogous to the population problem (2.8), it finds parameter values for α and β through the inverse step (3.2) such that the value of coefficient $\widehat{\gamma}(\alpha, \tau)$ on $\mathcal{Z}$ in the ordinary quantile regression step (3.1) is driven as close to zero as possible. This estimator is consistent and asymptotically normal under appropriate regularity and identification conditions:

$$\sqrt{n}(\widehat{\theta}(\tau) - \theta(\tau)) \to_d N(0, \Omega_\theta), \tag{3.4}$$

for Ω_θ specified below. This asymptotic distribution can be used for conducting direct inference on the parameter of interest using standard Wald procedures.

In addition, we can base inference on the "dual" Wald statistic $W_n(\alpha)$ for testing whether the coefficients on the instruments are zero (i.e. whether $\gamma(\alpha, \tau) = 0$). When $\alpha = \alpha(\tau)$, $W_n(\alpha)$ is asymptotically chi-squared with $\dim(\gamma)$ degrees of freedom:

$$W_n(\alpha(\tau)) \to_d \chi^2(\dim(\gamma)) \tag{3.5}$$

Thus, a valid confidence region for $\alpha(\tau)$ can also be based on the inversion of this dual Wald statistic:

$$CR_p[\alpha(\tau)] := \{\alpha : W_n(\alpha) < c_p\} \text{ contains } \alpha(\tau) \text{ with probability approaching } p, \tag{3.6}$$

where c_p is the p-percentile of a $\chi^2(\dim(\gamma))$ distribution. This dual approach is valid under weaker assumptions than the direct approach; in particular, it is robust to weak or partial identification of $\alpha(\tau)$. Section 3.5 discusses the properties of the dual procedure in more detail.

For a given probability index τ of interest, the estimator may be computed in practice as follows:

1. Define a suitable set of values $\{\alpha_j, j = 1, ..., J\}$, and run the ordinary τ-quantile regression of $Y_i - D_i'\alpha_j$ on X_i and $\mathcal{Z}_i$ to obtain coefficients $\widehat{\beta}(\alpha_j, \tau)$ and $\widehat{\gamma}(\alpha_j, \tau)$.

2. Save the inverse of the variance-covariance matrix of $\widehat{\gamma}(\alpha_j, \tau)$, which is readily available in any common implementation of the ordinary QR, to use as $\widehat{A}(\alpha_j)$ in $W_n(\alpha_j)$. Then $W_n(\alpha_j)$ becomes a Wald or F-statistic for testing $\gamma(\alpha_j, \tau) = 0$, depending on the naming convention.

3. Choose $\widehat{\alpha}(\tau)$ as a value among $\{\alpha_j, j = 1, ..., J\}$ that minimizes $W_n(\alpha)$. The estimate of $\beta(\tau)$ is then given by $\widehat{\beta}(\widehat{\alpha}(\tau), \tau)$.

4. Direct inference on $\alpha(\tau)$ may be conducted using the variance formula for Ω_θ provided below. Dual confidence regions for $\alpha(\tau)$, CR_p, may be computed as $CR_p[\alpha(\tau)] = \{\alpha_j : W_n(\alpha_j) < c_p\}$, and its upper and lower bounds may be used as end-points of a confidence interval for $\alpha(\tau)$.

3.2. Computational Complexity and Implementation

One of the most appealing features of the IVQR and associated dual inference confidence region is that both may be computed using the output from the conventional QR using any modern software. Portnoy and Koenker (1997) show that ordinary QR can be computed in polynomial stochastic time $O_p\left(\dim(\beta, \gamma)^3 \times n^{1+\delta}\right)$ using interior point algorithms with preprocessing,[3] so the above IVQR procedure has computational complexity of $O_p((1/\epsilon)^{\dim(\alpha)} \times \dim(\beta, \gamma)^3 \times n^{1+\delta})$ for a desired level of accuracy ϵ and some $\delta > 0$. Since we need $\epsilon \propto 1/n^a$, $a > 1/2$, and it suffices to have $a = 1/2 + \delta'$, for some small $\delta' > 0$, the proposed algorithm has computational complexity

$$O_p\left(n^{(1/2+\delta')\dim(\alpha)} \times \dim(\beta, \gamma)^3 \times n^{1+\delta}\right)$$

[3] In contrast, simplex procedures will have running time of $O_p\left(n^{\dim(\beta,\gamma)}\right)$.

that is polynomial in the sample size n and in the dimension of (β, γ), but is not polynomial in the dimension of α. Thus, the procedure will be computationally attractive and work well when the number of exogenous variables, $\dim(\beta)$, is possibly large, but the number of endogenous variables, $\dim(\alpha)$, is small. This situation is certainly the most common case prevalent in econometric specifications, where typically $\dim(\alpha) = 1$ or 2 and $\dim(\beta)$ varies from 1 to 50 or more. In fact, due to its practical properties, the estimator has already been applied in empirical analysis by Hausman and Sidak (2004), Januszewski (2004), and Chernozhukov and Hansen (2004a), and this paper also presents two additional empirical applications.

There are other approaches that one could adopt for estimation of the model defined in Section 2.1. For example, an immediate approach is the method of moments approach (MM) that attempts to minimize $\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1(Y_i \leq D_i'\alpha + X_i'\beta) - \tau)(X_i', \mathcal{Z}_i')'V_i\|$ over α and β. Another example is the estimator of Sakata (2001) which is an elegant maximum likelihood type estimator based on the absolute deviation.[4] In contrast to the IVQR approach, these alternative approaches involve highly non-convex, multi-modal, and non-smooth objective functions over many parameters, which poses a serious computational challenge. Implementation of extremum estimators with non-smooth and, more importantly, non-convex objective functions generally requires non-convex searches over parameter sets of dimension $K = \dim(\beta) + \dim(\alpha)$, which will be quite large in many cases due to the high-dimension of β. Thus, the IVQR approach will have an advantage when $\dim(\alpha)$ is small, as is the case in many applications. When $\dim(\alpha)$ is high, both the IVQR approach and the MM approach become difficult to implement. In such settings one could use the quasi-Bayesian methods for MM developed in Chernozhukov and Hong (2003). This approach computes estimates and confidence intervals using a quasi-posterior defined as the exponent of the MM function specified above.

4. Asymptotic Distribution Theory

4.1. Assumptions

To state the assumptions, define the population objective function as

$$Q(\tau, \alpha, \beta, \gamma) := E\left[\rho_\tau(Y_i - D_i'\alpha - X_i'\beta - \mathcal{Z}_i'\gamma)V_i\right], \tag{4.1}$$

and let

$$(\beta(\alpha, \tau), \gamma(\alpha, \tau)) := \arg\min_{\beta, \gamma} \; Q(\tau, \alpha, \beta, \gamma). \tag{4.2}$$

[4]In our notation, the estimator solves the following program:

$$\max_{\alpha,\beta}\min_{\gamma,\delta}\left[\sum_{i=1}^{n}|Y_i - D_i'\alpha - X_i'\beta - \mathcal{Z}_i'\gamma - X_i'\delta| / \sum_{i=1}^{n}|Y_i - D_i'\alpha - X_i'\beta|\right].$$

Define the parameter space $\Theta = \mathcal{A} \times \mathcal{B}$ as a compact convex set such that $\mathcal{B}$ contains the population value $\beta(\alpha, \tau)$ for each $\alpha \in \mathcal{A}$ in its interior, so that the parameter-on-the-boundary problem does not arise. We assume the data were generated by the model defined in Section 2 and impose the following additional assumptions:

R1 (Y_i, D_i, X_i, Z_i) are iid defined on the probability space (Ω, F, P) and have compact support.
R2 For the given τ, $(\alpha(\tau), \beta(\tau))$ is in the interior of the specified set Θ.
R3 Density $f_Y(Y|X, D, Z)$ is bounded by a constant $\bar{f}$ a.s.
R4 $\partial E\left[1(Y < D'\alpha + X'\beta + \mathcal{Z}'\gamma)\Psi\right]/\partial(\beta', \gamma')$ has full rank at each α in $\mathcal{A}$, for $\Psi = V_i(Z_i', X_i')'$

The compactness conditions in R1 and R2 simplify the analysis. The bounded density in R3 and compactness condition in R1 are sufficient for the Jacobian matrix in R4 to be well-defined. The full rank condition in R4 and iid sampling suffice for the estimates $(\widehat{\beta}(\alpha, \tau), \widehat{\gamma}(\alpha, \tau))$ to be asymptotically normal and are sufficient for implementing the dual inference. It is important to note that R1-R4 do not impose any conditions on the relation between D and Z; that is, unlike the direct inference procedure, the dual procedure will be valid when identification is weak or fails partially or completely.

Stronger additional conditions are imposed for implementing the direct inference.

R5 $\partial E\left[1(Y < D'\alpha + X'\beta)\Psi)'\right]/\partial(\alpha', \beta')$ has full rank at $(\alpha(\tau)', \beta(\tau)')'$.
R6 The function $(\alpha, \beta) \mapsto E\left[\{\tau - 1(Y < D'\alpha + X'\beta)\}\Psi\right]$ is one-to-one over Θ.

The imposition of R1-R6 is sufficient for identification and asymptotic normality of the IVQR estimator, both of which are necessary for the validity of the direct inference approach. These assumptions considerably strengthen the conditions R1-R4 by imposing restrictions on the relationship between D and Z. The dual approach does not require these assumptions for its validity. Hence, the dual approach is robust to the violation of either R5 or R6.

To further comment on the nature of correlation between Z and D required by R5, note that by R1 and R3 we have that

$$\partial E\left[1(Y < D'\alpha + X'\beta)\Psi)'\right]/\partial(\alpha', \beta') = E[f_\epsilon(0|X, D, Z)V_i(\mathcal{Z}_i', X_i)'(D_i', X_i')].$$

Hence, if we set $V_i = 1$ for simplicity, we see that the Jacobian in R5 takes a form of density-weighted covariation matrix between D and Z, and R5 requires that this matrix has full rank. R6 imposes that global identifiability must hold; hence, the impact of Z should be rich enough to guarantee that the moment equations are solved uniquely. These assumptions are required to carry out direct inference but are not required in the dual approach. Thus, discrepancies between the dual approach and the direct approach should be indicative of situations where R5 and R6 do not hold.

Before proceeding to the asymptotic results, we provide some sufficient and more primitive conditions for the global identifiability condition R6. A set of conditions that suffices for both R5 and R6 is as follows:

R5' $\partial E\left[1(Y < D'\alpha + X'\beta)\Psi\right]/\partial(\alpha', \beta')$ has full rank at each (α, β) in Θ.
R6' The image of Θ under $(\alpha, \beta) \mapsto E\left[\{\tau - 1(Y < D'\alpha + X'\beta)\}\Psi\right]$ is simply-connected.

R5' ensures that the mapping $(\alpha, \beta) \mapsto E\left[\{\tau - 1(Y < D'\alpha + X'\beta)\}\Psi\right]$ is locally one-to-one everywhere. The simple-connectivity condition R6' curbs somewhat the non-linearity of the mapping and implies a global one-to-one relationship by a Plastock-Hadamard type result, cf. Ambrosetti and Prodi (1995). This fact and equations (2.4) and (2.5) imply that the solution of the equation

$$E\left[\{\tau - 1(Y < D'\alpha + X'\beta)\}\Psi\right] = 0$$

is unique and is given by $(\alpha(\tau), \beta(\tau))$.

Other sufficient and more primitive conditions for R5 and R6 also result through an application of Theorem 2 in Mas-Colell (1979). Let Θ' be a convex compact set that contains Θ and that has a smooth boundary $\partial\Theta'$. Then the following conditions imply R5' and R6' and hence R5 and R6.

R5* $\partial E\left[1(Y < D'\alpha + X'\beta)\Psi\right]/\partial(\alpha', \beta')$ has a positive determinant at each (α, β) in Θ'.
R6* $\partial E\left[1(Y < D'\alpha + X'\beta)\Psi\right]/\partial(\alpha', \beta')$ is positive quasi-definite along the boundary $\partial\Theta'$ in the sense defined by Mas-Colell (1979).

Note that in the exogenous model, we can set $\mathcal{Z}_i = D_i$, and these conditions will be trivially satisfied.

4.2. **Asymptotic Properties of the Dual Inference**

We first state the formal results for dual inference, because they are the simplest to explain. Under the conditions R1-R4, as $n \to \infty$, uniformly in $\alpha \in \mathcal{A}$

$$\sqrt{n}(\widehat{\vartheta}(\alpha, \tau) - \vartheta(\alpha, \tau)) = -J_\vartheta^{-1}(\alpha) \cdot n^{-1/2} \cdot \sum_{i=1}^{n} s_i(\alpha) + o_p(1), \tag{4.3}$$

$$s_i(\alpha) = \left[\tau - 1\left(\epsilon_i(\alpha) < 0\right)\right]\Psi_i, \Psi_i = V_i\{X_i', \mathcal{Z}_i'\}', \tag{4.4}$$

$$\epsilon_i(\alpha) = Y_i - D_i\alpha - X_i'\beta(\alpha, \tau) - \mathcal{Z}_i'\gamma(\alpha, \tau), \tag{4.5}$$

$$J_\vartheta(\alpha) = E\left[f_{\epsilon(\alpha)}(0|Z, X)\Psi\Psi'/V\right]. \tag{4.6}$$

(4.3)-(4.6) follow by adopting standard arguments for the quantile regression process. The difference here is that we have a process in α, whereas we usually have a process over τ. Hence for each α

$$\sqrt{n}\left(\widehat{\vartheta}(\alpha, \tau) - \vartheta(\alpha, \tau)\right) \to_d N\left(0, \Omega_\vartheta[\alpha]\right), \tag{4.7}$$

$$\Omega_\vartheta[\alpha] = J_\vartheta^{-1}[\alpha]S[\alpha]J_\vartheta^{-1}[\alpha], \quad S[\alpha] = E[s_i(\alpha)s_i(\alpha)']. \tag{4.8}$$

The statistic for testing $\gamma(\alpha,\tau)=0$ is given by the Wald statistic $W_n(\alpha)=n[\widehat{\gamma}(\alpha,\tau)']\widehat{\Omega}_\vartheta[\alpha][\widehat{\gamma}(\alpha,\tau)]$, where $\widehat{\Omega}_\vartheta[\alpha]=\Omega_\vartheta[\alpha]+o_p(1)$ is any standard consistent estimate of the asymptotic variance (4.8) of the ordinary QR. Therefore, when $\alpha=\alpha(\tau)$

$$W_n[\alpha(\tau)] \rightarrow_d \chi^2(\dim(\gamma)), \tag{4.9}$$

and for the confidence region $CR_p[\alpha(\tau)]:=\{\alpha\in\mathcal{A}:W_n(\alpha)<c_p\}$, where $P\{\chi^2(\dim(\gamma))<c_p\}=p$,

$$P\{\alpha(\tau)\in CR_p[\alpha(\tau)]\}=P\{W_n[\alpha(\tau)]<c_p\}\rightarrow p. \tag{4.10}$$

Proposition 1. *Under conditions R1-R4, the results (4.3)-(4.10) are true.*

Comment 1. Unlike direct inference, the dual inference requires only assumptions R1-R4 to hold. The results for dual inference are also straightforward to extend in various direction. In particular, one can note that the preliminary estimation of weights V_i and instruments $\mathcal{Z}_i$ will not affect (4.9) or even (4.7) as long as α is in a root-n neighborhood of $\alpha(\tau)$. Additional regularity conditions on the estimates of weights and instruments that must be imposed can be found in Andrews (1994).

4.3. **Asymptotic Properties of the Direct Inference**

The following proposition presents the asymptotic properties of the direct approach.

Proposition 2. *In the specified model under conditions R1-R4 and conditions R5-R6, sufficient conditions for which are either R5'-R6' or R5*-R6*,*

$$\sqrt{n}\left(\widehat{\theta}(\tau)-\theta(\tau)\right)\rightarrow_d N(0,\Omega_\theta),\quad \Omega_\theta=(K',L')'S(K',L'), \tag{4.11}$$

where, for $\Psi=V\cdot[X',\mathcal{Z}']'$ *and* $\epsilon=Y-D'\alpha(\tau)-X'\beta(\tau)$, $S=\tau(1-\tau)E\left[\Psi\Psi'\right]$, $K=(J_\alpha' H J_\alpha)^{-1}J_\alpha' H$, $H=\bar{J}_\gamma' A[\alpha(\tau)]\bar{J}_\gamma$, $L=\bar{J}_\beta M$, $M=I_{k+r}-J_\alpha K$, $J_\alpha=E[f_\epsilon(0|X,Z,D)\Psi D']$, *and* $[\bar{J}_\beta',\bar{J}_\gamma']'$ *is a partition of* $J_\vartheta^{-1}:=\left(E\left[f_\epsilon(0|X,Z)\Psi\Psi'/V\right]\right)^{-1}$ *such that* $\bar{J}_\beta$ *is a* $\dim(\beta)\times\dim(\beta,\gamma)$ *matrix and* $\bar{J}_\gamma$ *is a* $\dim(\gamma)\times\dim(\beta,\gamma)$ *matrix.*

Corollary 1. *When* $\dim(\gamma)=\dim(\alpha)$, *the choice of* $A(\tau)$ *does not affect asymptotic variance, and the joint asymptotic variance of* $\widehat{\alpha}(\tau)$ *and* $\widehat{\beta}(\tau)$ *will generally have the simple form*

$$\Omega_\theta=J_\theta^{-1}S(J_\theta')^{-1}$$

for S *defined above and* $J_\theta=E[f_\epsilon(0|X,Z,D)\Psi[D',X']]$.

Corollary 2. *When* $\dim(\gamma)>dim(\alpha)$, *the choice of the weighting matrix* $A(\alpha)$ *in the objective function* $W_n(\alpha)$ *generally matters. A natural choice for* $A(\alpha)$ *is given by* $A(\alpha)=([\Omega_\vartheta[\alpha]_{22})^{-1}$ *which corresponds to the inverse of the covariance matrix of* $\sqrt{n}(\widehat{\gamma}(\alpha,\tau)-\gamma(\alpha,\tau))$. *Noting that* $A(\alpha)$ *is equal to* $(\bar{J}_\gamma S\bar{J}_\gamma')^{-1}$ *at* $\alpha(\tau)$, *it follows that the asymptotic variance of* $\sqrt{n}(\widehat{\alpha}(\tau)-\alpha(\tau))$ *equals*

$$\Omega_\alpha=(J_\alpha'\bar{J}_\gamma'(\bar{J}_\gamma S\bar{J}_\gamma')^{-1}\bar{J}_\gamma J_\alpha)^{-1}.$$

Corollary 3. *The efficient score for* $(\alpha(\tau), \beta(\tau))$ *is given by* $\frac{1}{\tau(1-\tau)}[\tau - 1(\epsilon \leq 0)]\Psi^*$*, where* $\Psi^* = V^* \cdot [X', \mathcal{Z}^{*\prime}]'$*,* $\mathcal{Z}^* := E[D \cdot v^*|Z, X]/V^*$*,* $v^* := f_\epsilon(0|D, Z, X)$*, and* $V^* = f_\epsilon(0|Z, X)$*. Thus, if* $\mathcal{Z} = \mathcal{Z}^*$ *and* $V = V^*$*, the asymptotic variance of* $(\widehat{\alpha}(\tau), \widehat{\beta}(\tau))$ *attains the efficiency bound* $\tau(1-\tau)E[\Psi^*\Psi^{*\prime}]^{-1}$*.*

Comment 2. Corollary 1 is especially convenient since the variance formula becomes simple once the instrument $\mathcal{Z}$ is collapsed to the same dimension as D. Corollary 3 shows how to construct the instrument $\mathcal{Z}$ and weight V such that the IVQR estimator achieves the efficiency bound in the sense defined by Amemiya (1977) as well as the semi-parametric efficiency bound in the sense of Bickel, Klaassen, Ritov, and Wellner (1993). Efficient estimation can be implemented in two steps. In the first step, IVQR is used to obtain residuals $\widehat{\epsilon}_i$. In the second step, the required weights V_i^* and instruments $\mathcal{Z}_i^*$ are estimated using nonparametric or parametric methods and are used in IVQR again. It can be shown that estimation of weights and instruments has no effect on the limit distribution of the estimators, provided additional regularity conditions, found e.g. in Andrews (1994), on the estimates of weights and instruments hold.

4.4. Estimating Variance and Jacobian Matrices

The components of the variance matrices that we need to estimate include J_ϑ, J_α, and S for direct inference and $J_\vartheta[\alpha]$ and $S[\alpha]$ for dual inference. Following Koenker's (1994) analysis for ordinary QR, the first set of components can be estimated as follows:

$$\widehat{S} = \frac{1}{n}\sum_{i=1}^{n} \widehat{s}_i \widehat{s}_i', \quad \widehat{J}_\alpha = \frac{1}{n}\sum_{i=1}^{n} [K(\widehat{\epsilon}_i/h)/h]\, \Psi_i D_i', \quad \widehat{J}_\vartheta = \frac{1}{n}\sum_{i=1}^{n} [K(\widehat{\epsilon}_i/h)/h]\, \Psi_i \Psi_i'/V_i,$$

where $\widehat{\epsilon}_i := Y_i - D_i'\widehat{\alpha}(\tau) - X_i'\widehat{\beta}(\tau)$, $\widehat{s}_i = [\tau - 1\,(\widehat{\epsilon}_i < 0)]\, \Psi_i$, $\Psi_i = V_i[Z_i', X_i']'$, h is a bandwidth chosen such that $h \to 0$ and $nh^2 \to \infty$, and $K(\cdot)$ is a kernel function. Specific choices of h are discussed in Koenker (1994). Similarly, the second set of estimates is given by

$$\widehat{S}[\alpha] = \frac{1}{n}\sum_{i=1}^{n} \widehat{s}_i[\alpha]\widehat{s}_i[\alpha]', \quad \widehat{J}_\vartheta[\alpha] = \frac{1}{n}\sum_{i=1}^{n} [K(\widehat{\epsilon}_i[\alpha]/h)/h]\, \Psi_i \Psi_i'/V_i$$

where $\widehat{\epsilon}_i[\alpha] := Y_i - D_i'\alpha - X_i'\widehat{\beta}(\tau, \alpha) - \mathcal{Z}_i'\widehat{\gamma}(\alpha, \tau)$ and $s_i[\alpha] = [\tau - 1\,(\widehat{\epsilon}_i[\alpha] < 0)]\, \Psi_i$,$\Psi_i = V_i[Z_i', X_i']'$, h is a bandwidth chosen such that $h \to 0$ and $nh^2 \to \infty$, and $K(\cdot)$ is a kernel function. The consistency properties of these estimators are standard and will not be discussed here.

5. Empirical Examples

In this section, we present two applications of the estimation and inference results derived in Section 3. The first application reports the results of a simple analysis of the demand for fish. This application makes use of a small sample and illustrates the potential differences between the direct and dual inference procedures. In the second example, we consider the effects of a job training program on earnings. In this case, the identification is quite strong, and we see small differences between the direct and dual inference procedures. The results here also demonstrate the bias in the conventional quantile regression under endogeneity. In particular, the conventional quantile regression estimates indicate that the effect of training is positive and significant across the entire outcome distribution, while the IVQR estimates indicate that the training impact is close to zero in the lower tail of the outcome distribution.

5.1. **Demand for Fish**

In this section, we present estimates of demand elasticities which may potentially vary with the level of demand. The data contain observations on price and quantity of fresh whiting sold in the Fulton fish market in New York over the five month period from December 2, 1991 to May 8, 1992. These data were used previously in Graddy (1995) to test for imperfect competition in the market. The price and quantity data are aggregated by day, with the price measured as the average daily price and the quantity as the total amount of fish sold that day. The total sample consists of 111 observations for the days in which the market was open over the sample period.

For the purposes of this illustration, we focus on a simple Cobb-Douglas random demand model with non-additive disturbance:

$$\ln(Q_p) = \alpha_0(U) + \alpha_1(U)\ln(p),$$

where Q_p is the quantity that would be demanded if the price were p, U is an unobservable affecting the level of demand normalized to follow $U(0,1)$, and $\alpha_1(U)$ is the random demand elasticity when the level of demand is U. A supply function $S_p = f(p, Z, \mathcal{U})$ describes how much producers would supply if the price were p, subject to other factors Z and unobserved disturbance $\mathcal{U}$. The factors Z affecting supply are assumed to be independent of demand disturbance U.

The observed quantity Y sold in the market is given in logs by

$$\begin{aligned} &\ln Y = \alpha_0(U) + \alpha_1(U)\ln P, \text{ where} \\ &U \text{ is independent of } Z, \end{aligned} \tag{5.1}$$

where P is the price picked by the market to equate supply and demand. That is, P satisfies $\alpha_0(U) + \alpha_1(U)\ln(P) = \ln f(P, Z, \mathcal{U})$, which implies the observed price depends on the demand disturbance U, i.e. that $P = \delta(Z, U, \mathcal{U})$ for some function δ.

As instruments Z, we consider two different variables capturing weather conditions at sea: *Stormy* is a dummy variable which indicates wave height greater than 4.5 feet and wind speed greater than 18 knots, and *Mixed* is a dummy variable indicating wave height greater than 3.8

feet and wind speed greater than 13 knots. These variables are plausible instruments since weather conditions at sea should influence the amount of fish that reaches the market but should not influence demand for the product.[5] Simple OLS regressions of the log of price on these instruments suggest they are correlated to price, yielding R^2 and F-statistics of 0.227 and 15.83 when both Stormy and Mixed are used as instruments and 0.160 and 20.69 when just Stormy is used. However, given the small sample, we may still expect identification to be weak, and weak identification is suggested by the results reported below.

Quantile regression (QR) and inverse quantile regression (IVQR) results for the .15, .25, .50, .75, and .85 quantiles are reported in columns (1)-(3) of Table 1 below. Column (1) reports the QR results, while columns (2) and (3) report IVQR results. Columns (2) and (3) differ in that only Stormy is used as an instrument in Column (2), while Stormy and Mixed are used as instruments in Column (3). For the τ^{th} quantile, the row labeled $\widehat{\alpha}(\tau)$ gives the QR or IVQR estimate of α. The row labeled "Wald Interval" contains the 95% confidence interval for $\widehat{\alpha}(\tau)$ constructed based on the asymptotic approximation, and for the IVQR estimates, the row labeled "Dual Interval" contains the 95% confidence bound constructed using the dual inference procedure outlined in Section 3.1. The computation of the IVQR estimator was conducted over the parameter space $\mathcal{A} = [-5, 5]$ using α_j equally spaced with a step size of 0.1.

The IVQR estimates exhibit considerable heterogeneity, ranging from -1.5 to -0.7 in column (2) and from -1.5 to -0.9 in column (3). The IVQR elasticities are also uniformly greater in magnitude than the "price effects" estimated by the ordinary QR, which we might anticipate given endogenous sampling resulting from the joint determination of price and quantity in the market. These differences are illustrated graphically in Figure 1. The left panel reports the QR results, and the right panel reports the IVQR results when both Stormy and Mixed are used as instruments. In the figure, we clearly see that the demand functions estimated by IVQR are steeper than those estimated by QR when plotted in log-price-log-quantity space, and that this translates directly into more curvature of the demand curve when plotted in the original price-quantity space. It is also important to note that the interpretation of IVQR and QR estimates is very different. IVQR estimates a structural demand model, while QR estimates the conditional quantiles of the equilibrium quantity variable as a function of the equilibrium price. It is no surprise that these estimates are different.

Interestingly, there are clear and large differences between the confidence intervals given by the two different inference procedures for IVQR. In particular, the confidence intervals based on the dual procedure which is robust to weak and partial identification are uniformly much wider than the intervals based on the direct inference procedure. For instance, the dual confidence region for the .85 quantile case contains the upper endpoint of the parameter space $\mathcal{A}$. In addition, it is worth

[5]More detailed arguments may be found in Graddy (1995).

Table 1. Results from Empirical Examples

	Example 1: Demand for Fish			Example 2: Returns to Training	
	QR	IVQR	IVQR	QR	IVQR
	(1)	(2)	(3)	(4)	(5)
$\widehat{\alpha}(.15)$	-0.53	-1.5	-1.5	1188	-200
Wald CI	(-1.24,0.16)	(-3.69,-0.69)	(-2.51,-0.49)	(553,1822)	(-1435,1035)
Dual CI		[-5.0,0.5)	(-3.2,0.1)		(-1300,1500)
$\widehat{\alpha}(.25)$	-0.40	-1.0	-1.4	2510	500
Wald CI	(-0.87,0.07)	(-2.51,0.51)	(-2.52,-0.28)	(1742,3278)	(-887,1887)
Dual CI		(-4.4,0.0)	(-3.1,0.1)		(-1000,2000)
$\widehat{\alpha}(.50)$	-0.41	-0.7	-0.9	4420	300
Wald CI	(-0.81,-0.01)	(-1.67,0.27)	(-1.82,0.02)	(3220,5621)	(-1589,2189)
Dual CI		(-3.0,0.6)	(-3.0,0.6)		(-1400,2700)
$\widehat{\alpha}(.75)$	-0.70	-1.2	-1.3	4678	2700
Wald CI	(-1.18,-0.22)	(-2.02,-0.38)	(-2.07,-0.53)	(2901,6455)	(-260,5660)
Dual CI		(-2.0,-0.1)	(-2.1,0.1)		(-400,5600)
$\widehat{\alpha}(.85)$	-0.81	-1.3	-1.1	4806	3200
Wald CI	(-1.24,-0.38)	(-2.10,-0.50)	(-1.82,-0.38)	(2751,6861)	(32, 6368)
Dual CI		(-2.0,5.0]	(-2.6,5.0]		(500,5800)

Notes: Columns (1)-(3) report results from estimation of the demand for fish, and columns (4) and (5) report results from estimation of the returns to training from the JTPA experiment. Columns (1) and (4) report conventional quantile regression results, and columns (2), (3), and (5) report instrumental quantile regression results. In column (2), one instrument, *Stormy*, is used, and in column (3), two instruments, *Stormy* and *Mixed* are used. Rows labeled $\alpha(\tau)$ for $\tau \in \{.15, .25, .50, .75, .85\}$ report point estimates, and the numbers in parentheses are confidence intervals.

noting that the confidence intervals obtained using two instrumental variables are generally shorter than the confidence intervals obtained using just one instrumental variable, suggesting an efficiency gain to using more instruments.

The construction and nature of the dual confidence bounds are further illustrated in Figures 2 and 3, which respectively plot the IVQR objective function $W_n(\alpha)$ over the parameter space in the two cases. α is plotted on the horizontal axis, and the vertical axis shows $W_n(\alpha)$. The horizontal line in each graph is the 95% critical value for the dual testing procedure, so all points lying below the horizontal line belong to the confidence region for $\alpha(\tau)$.

These graphs display a number of interesting features. It is apparent that the objective function, while having numerous local minima, has a distinct minimum over $\mathcal{A}$ in all cases. The objective functions, and hence dual confidence regions, are generally well-behaved in the middle of the distribution and become more erratic as one moves toward the tails of the distribution. It is also clear that the dual confidence regions are not connected in many cases.

This simple example clearly illustrates the potential differences between the direct and dual inference procedures. It also provides an example of an application of the methods of this paper to demand analysis where the elasticity of demand is potentially heterogeneous. The next example illustrates the use of the estimator in a setting with a considerably larger sample and where identification is much stronger, showing that in this setting the two inference procedures produce similar results. The results also demonstrate the interesting insights that may be gained through quantile analysis and the importance of accounting for endogeneity in such studies.

5.2. **The Returns to Training**

The impact of job training programs on the earnings of trainees, especially those with low income, is of great interest to both policy makers and academic economists, but evaluating the causal effect of training programs on earnings is difficult due to the self-selection of treatment status. However, data available from a randomized training experiment conducted under the Job Training Partnership Act (JTPA) provides a mechanism for addressing this issue. In the experiment, people were randomly assigned the offer of JTPA training services, but because people were able to refuse to participate, the actual treatment receipt was self-selected. Of those offered treatment, only 60 percent participated in the training. There was also a small number of individuals from the control group who received training. The random assignment of the training offer provides a plausible instrument for a person's actual training status. Abadie, Angrist, and Imbens (2002), who previously used this data to examine the impact of job training on earnings,[6] provide more detailed information regarding data collection procedures, sample selection criteria, and institutional details of the JTPA along with additional facts and discussion about the JTPA training experiment. In this example, we limit the analysis to the sample of adult males.

To capture the effects of training on earnings, we estimate a structural quantile model of the form

$$Y = D\alpha(U) + X'\beta(U), \quad U \sim U(0,1), \text{ given } \ Z \text{ and } X,$$

where D indicates training status and is instrumented for by assignment to the treatment group, the outcomes Y are earnings, X is a vector of covariates, Z is a dummy variable indicating assignment to the treatment group, and U is an unobservable affecting earnings.

[6]They use a different modeling framework and estimator that estimates the treatment effect for the sub-population of "compliers".

The data consist of 5,102 observations with data on earnings, training and assignment status, and other individual characteristics. Earnings are measured as total earnings over the 30 month period following the assignment into the treatment or control group, and average earnings in the sample are $19,147. The vector of controls, X, includes dummies for black and Hispanic persons, a dummy indicating high-school graduates and GED holders, five age-group dummies, a marital status dummy, a dummy indicating whether the applicant worked 12 or more weeks in the 12 months prior to the assignment, a dummy signifying that earnings data are from a second follow-up survey, and dummies for the recommended service strategy.[7] For brevity, we only report results for estimates of the key parameter, $\alpha(\tau)$, which represents the impact of the training program on earnings.

Since assignment to the treatment or control group was random, it provides a natural instrument Z. The instrument is useful for identification since it is highly correlated to the actual training state D. The partial R^2 of a regression of training status on assignment to the treatment group, controlling for the other covariates, is .609, and the first-stage F-statistic is 2,673. This strong correlation suggests that weak identification should not be a problem in this case, so the direct and dual inference procedures should yield similar results.

As in the previous example, estimation results for the .15, .25, .50, .75, and .85 quantiles are reported in columns (4)-(5) of Table 1. Column (4) reports the QR results, while column (5) reports the IVQR results. For the τ^{th} quantile, the row labeled $\widehat{\alpha}(\tau)$ gives the QR or IVQR estimate of α. The row labeled "Wald Interval" contains the 95% confidence interval for $\widehat{\alpha}(\tau)$ constructed based on the asymptotic approximation, and for the IVQR estimates, the row labeled "Dual Interval" contains the 95% confidence bound constructed using the dual inference procedure outlined in Section 3.1. The computation of IVQR was conducted over the parameter space $\mathcal{A} = [-2500, 7500]$ using α_j equally spaced with a step size of 100. In addition, the results are presented graphically in Figure 4.

The conventional QR results, which fail to account for the selection into the treatment state, are uniformly positive and significantly different from 0. They indicate that the training program had a relatively large impact on the earnings of participants, and that this impact is increasing in the quantile index. However, given that people were able to decide whether or not to participate in training following the initial random assignment, it seems likely that these estimates would be upward biased for the actual effect of training on earnings. This suspicion is confirmed by the IVQR estimates, which account for the endogeneity of training status and are uniformly smaller than the corresponding QR estimates. This difference is most apparent in the low and middle earning quantiles. In the low quantiles, QR suggests a moderate positive and significant effect of training on earning quantiles; however, the IVQR estimates are quite low and, while imprecise, not significantly

[7]The recommended service strategy was broken into three categories: classroom training, on-the-job training and/or job search assistance, and other forms of training.

different from 0. The difference in the estimates becomes more apparent when we consider the percentage impact of the training program, which is presented in the right hand column of Figure 4.[8] Here, the QR estimates imply a large percentage increase in earnings in the low earning quantiles, starting at 139% for $\tau = .15$, which declines as one moves to the upper quantiles of the conditional earnings distribution, though the impact remains large even in the center of the distribution at $\tau = .50$, where the implied effect is a 35% increase in earnings due to training. The IVQR estimates on the other hand are quite stable, varying between -13% and 14%, and with the exception of $\tau = .25$ are all below 10%.

Unlike the case considered above, we do not find large differences between the direct and dual inference procedures for IVQR in this case. The similarity between the two approaches is not unexpected due to the strong correlation between the instrument and endogenous regressor. The close agreement here further suggests that not much is lost by considering the dual procedure in cases where identification is strong. It also provides further support for the argument that the differences detected in the previous section are due to weak identification. Given the robustness of the dual procedure to the presence of weak instruments and its simple computation, it seems that this inference procedure will be preferable to the standard procedure in many cases.

The dual confidence bounds are further illustrated in Figure 5, which plots the IVQR objective function $W_n(\alpha)$ over the parameter space $\mathcal{A}$. α is plotted on the horizontal axis, and the vertical axis shows $W_n(\alpha)$. The horizontal line in each graph is the 95% critical value for the dual testing procedure, so all points lying below the horizontal line belong to the confidence region for $\alpha(\tau)$. The graphs in Figure 3 differ markedly from those in Figures 1 and 2. In particular, all of the objective functions, and hence confidence regions, in Figure 3 look remarkably well-behaved. The objective functions appear to be reasonably smooth, and the confidence intervals are all connected and clearly bounded within the parameter space considered.

6. Conclusion

In this paper, we propose an estimation approach, the inverse quantile regression (IVQR), that appropriately modifies the conventional quantile regression (QR) and recovers quantile-specific covariate effects in an instrumental variables model defined by $Y = D'\alpha(U)$ where U is independent of a set of instrumental variables Z. The IVQR estimator is appealing for estimation in this model since it can be computed through a series of conventional quantile regression steps and so will be computationally convenient in many cases encountered in practice. We derive the asymptotic properties of the estimator under suitable conditions. In addition, we demonstrate that the estimation

[8] The percentage impact is for changing from $D = 0$ to $D = 1$, i.e. from the non-training to the training state. All other covariates were evaluated at their sample means.

procedure leads to a testing procedure which will be robust to the presence of weak instruments and that this inference procedure results naturally from the IVQR algorithm and so is simple to implement in practice.

We then illustrate the use of the proposed estimator and testing procedure through two brief empirical examples. In the first example, we examine a simple demand model in a small sample with relatively weak instruments. In this case, we find that the conventional QR estimate of the elasticity of demand appears to be upward biased as would be expected due to the joint determination of price and quantity by supply and demand. In addition, we find that there are large differences between the direct inference procedure and the dual inference procedure which is robust to weak instruments. In the second example, we look at the impact of a job training program on earnings. In this case, we instrument for training status with random assignment to the training program which is very highly correlated to actual receipt of training. In this case, there is essentially no difference between the direct inference procedure and the dual procedure which is robust to weak instruments. In addition, there is strong evidence of endogeneity of training status resulting in substantial bias to the conventional QR estimator. This bias is especially pronounced in the lower tail of the earnings distribution where QR suggests a significant and positive effect of training on earnings, while the IVQR estimates are insignificant and small in magnitude.

7. Appendix

7.1. Proof of Proposition 1

The result (4.3) follows by adopting standard arguments for quantile regression processes, for instance those of Gutenbrunner and Jurečková (1992). The rest of the stated conclusions (4.4)-(4.10) follow by the Slutsky Lemma.

7.2. Proof of Proposition 2

We use standard definitions and notation from empirical process theory, as e.g. van der Vaart and Wellner (1996) and van der Vaart (1998). For $W := (Y, D, X, Z)$, define the maps

$$f \mapsto \mathbb{E}_n\left[f(W)\right] := \frac{1}{n}\sum_{i=1}^{n} f(W_i), \quad f \mapsto \mathbb{G}_n\left[f(W)\right] := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(f(W_i) - \mathbb{E}\left[f(W_i)\right]\right), \tag{7.1}$$

where we use E to denote the usual expectation and $\mathbb{E}$ to denote expectation evaluated at an estimated function $\widehat{f}$: $\mathbb{E}[\widehat{f}(W_i)] := \left(E\left[f(W_i)\right]\right)_{f=\widehat{f}}$.

For convenience we collect important definitions below. Let $\vartheta := (\beta, \gamma)$ and $\vartheta(\tau) := (\beta(\tau), 0)$. Define

$$f(W, \alpha, \vartheta) := \{\tau - 1(Y \leq D'\alpha + X'\beta + \mathcal{Z}'\gamma)\}\Psi, \tag{7.2}$$

where $\Psi := V \cdot (X', \ \mathcal{Z}')'$. Define for $\rho_\tau(u) = (\tau - 1(u < 0))u$

$$g(W, \alpha, \vartheta) := \rho_\tau(Y - D'\alpha - X'\beta - \mathcal{Z}'\gamma)V. \tag{7.3}$$

Define

$$Q_n(\alpha, \vartheta) := \mathbb{E}_n\left[g(W, \alpha, \vartheta)\right], \quad Q(\alpha, \vartheta) := E\left[g(W, \alpha, \vartheta)\right], \tag{7.4}$$

and

$$\widehat{\vartheta}(\alpha, \tau) := (\widehat{\beta}(\alpha, \tau), \widehat{\gamma}(\alpha, \tau)) := \ \arg\inf_{\vartheta \in \mathbb{R}^{\dim(\beta,\gamma)}} Q_n(\alpha, \vartheta), \tag{7.5}$$

$$\vartheta(\alpha, \tau) := (\beta(\alpha, \tau), \gamma(\alpha, \tau)) := \ \arg\inf_{\vartheta \in \mathbb{R}^{\dim(\beta,\gamma)}} Q\ (\alpha, \vartheta), \tag{7.6}$$

$$W_n[\alpha] := \widehat{\gamma}(\alpha, \tau)'\widehat{A}(\alpha)\widehat{\gamma}(\alpha, \tau), \quad W[\alpha] := \gamma(\alpha, \tau)'A(\alpha)\gamma(\alpha, \tau), \tag{7.7}$$

$$\widehat{\alpha}(\tau) \ := \arg\inf_{\alpha \in \mathcal{A}} W_n[\alpha], \quad \alpha^* := \arg\inf_{\alpha \in \mathcal{A}} W[\alpha], \tag{7.8}$$

$$\widehat{\beta}(\tau) := \widehat{\beta}(\widehat{\alpha}(\tau), \tau), \quad \beta^* := \beta(\alpha^*, \tau), \quad \widehat{\gamma}(\tau) := \widehat{\gamma}(\widehat{\alpha}(\tau), \tau), \quad \gamma^* := \gamma(\alpha^*, \tau). \tag{7.9}$$

Step 1 (Identification) We show that $\vartheta(\tau) = (\alpha(\tau)', \beta(\tau)')$ uniquely solves the limit problem.

First, by R6, the mapping $(\alpha, \beta) \mapsto E\left[\{\tau - 1(Y \leq D'\alpha + X'\beta)\}\Psi\right]$ is one-to-one over $\mathcal{A} \times \mathcal{B}$. By equation (2.5), we have that $\vartheta(\tau) = (\alpha(\tau)', \beta(\tau)')'$ solves the equation $E\left[\{\tau - 1(Y \leq D'\alpha + X'\beta)\}\Psi\right] = 0$, and it is thus the only solution over $\mathcal{A} \times \mathcal{B}$.

Second, we need to show that R5'-R6' and R5*-R6* suffice for R6. Sufficiency of R5'-R6' follows by a variant of Hadamard-Cacciopoli theorem for general metric spaces, cf. Theorem 1.8 in Ambrosetti and Prodi (1995). Sufficiency of R5*-R6* follows by Theorem 2 in Mas-Colell (1979).

Third, we have that the true parameters $(\alpha,\beta) = (\alpha(\tau),\beta(\tau))$ uniquely solve the equation

$$E\left[\{\tau - 1(Y \le D'\alpha + X'\beta + \mathcal{Z}'0)\}\Psi\right] = 0 \tag{7.10}$$

over $\mathcal{A}\times\mathcal{B}$. By R4 and by convexity in ϑ of the limit optimization problem for each α, $\vartheta(\alpha,\tau)$ uniquely solves the equation:

$$E\left[(\tau - 1\{Y \le D'\alpha + X'\beta(\alpha,\tau) + \mathcal{Z}'\gamma(\alpha,\tau)\})\,\Psi\right] = 0. \tag{7.11}$$

By construction of $\mathcal{A}\times\mathcal{B}$ we know that $\beta(\alpha,\tau)$ is in the interior of $\mathcal{B}$ for each $\alpha\in\mathcal{A}$. We need to find $\alpha^*\in\mathcal{A}$ such that this equation holds and the norm of $\gamma(\alpha^*,\tau)$ is minimal. $\alpha^* = \alpha(\tau)$ makes $\gamma(\alpha^*,\tau) = 0$ by equation (2.5). Thus $\alpha^* = \alpha(\tau)$ is a solution; by the preceding argument it is unique and $\beta(\alpha^*(\tau),\tau) = \beta(\tau)$.

Step 2. (Consistency) One consequence of Proposition 1, namely of equation (4.3), is that

$$\sup_{\alpha\in\mathcal{A}}\left\|\widehat{\vartheta}(\alpha,\tau) - \vartheta(\alpha,\tau)\right\| \to_p 0 \text{ i.e. } \sup_{\alpha\in\mathcal{A}}\left\|\widehat{\gamma}(\alpha,\tau) - \gamma(\alpha,\tau)\right\| \to_p 0, \tag{7.12}$$

which implies $\sup_{\alpha\in\mathcal{A}}\left|W_n(\alpha) - W(\alpha)\|\right| \xrightarrow{p} 0$, where $W(\alpha)$ is continuous in α over $\mathcal{A}$. It therefore follows by the standard consistency argument for extremum estimators that $\widehat{\alpha}(\tau)\to_p \alpha(\tau)$, and then by (7.12) that for any $\alpha_n \to_p \alpha(\tau)$, $\widehat{\beta}(\alpha_n,\tau) \to_p \beta(\alpha(\tau),\tau) = \beta(\tau)$ and $\widehat{\gamma}(\alpha_n,\tau)\to_p \gamma(\alpha(\tau),\tau) = \gamma(\tau) = 0$. Hence we also have that

$$\widehat{\vartheta}(\alpha_n,\tau)\to_p \vartheta(\alpha(\tau),\tau) \text{ for any } \alpha_n\to_p \alpha(\tau). \tag{7.13}$$

Note that above we have used that $\vartheta(\alpha,\tau)$ is continuous in α, which is verified by the implicit function theorem applied to equation (7.11).

Step 3. (Asymptotics) Let α_n be in a small ball centered at $\alpha(\tau)$. By the computational properties of the quantile regression estimator $\widehat{\vartheta}(\alpha_n,\tau)$ established in Theorem 3.3 in Koenker and Bassett (1978),

$$O(1/\sqrt{n}) = \sqrt{n}\mathbb{E}_n[f(W,\alpha_n,\widehat{\vartheta}(\alpha_n,\tau))]. \tag{7.14}$$

The functional class $\{f(W,\alpha,\vartheta),(\alpha,\vartheta)\in\mathcal{A}\times\mathcal{B}\times\mathcal{G}\}$ is Donsker for any compact sets $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{G}$, because this class is a product of a VC subgraph class and a bounded random vector. Hence the following expansion of the rhs of (7.14) is valid for any $\alpha_n\to_p \alpha(\tau)$

$$\begin{aligned} O(1/\sqrt{n}) &= \mathbb{G}_n\left[f\left(W,\alpha_n,\widehat{\vartheta}(\alpha(\tau),\tau)\right)\right] + \sqrt{n}\mathbb{E}\left[f\left(W,\alpha_n,\widehat{\vartheta}(\alpha_n,\tau)\right)\right] \\ &= \mathbb{G}_n\left[f\left(W,\alpha(\tau),\vartheta(\alpha(\tau),\tau)\right)\right] + \sqrt{n}\mathbb{E}\left[f\left(W,\alpha_n,\widehat{\vartheta}(\alpha_n,\tau)\right)\right] + o_p(1). \end{aligned} \tag{7.15}$$

Expanding the very last element further, by R4

$$
\begin{aligned}
O(1/\sqrt{n}) &= \mathbb{G}_n\left[f(W,\alpha(\tau),\vartheta(\tau))\right] + o_p(1) \\
&\quad + (J_\vartheta + o_p(1))\sqrt{n}(\widehat{\vartheta}(\alpha_n,\tau) - \vartheta(\tau)) + (J_\alpha + o_p(1))\sqrt{n}(\alpha_n - \alpha(\tau)),
\end{aligned} \tag{7.16}
$$

where by R1 and R3

$$
\begin{aligned}
J_\vartheta &= \frac{\partial}{\partial(\beta',\gamma')}\; E\left[\varphi_\tau(Y - D'\alpha(\tau) - X'\beta - \mathcal{Z}'\gamma)\Psi\right]\Big|_{(\gamma,\beta)=(0,\beta(\tau))} \\
&= E\left[f_\epsilon(0|X,Z)\Psi\Psi'/V\right], \\
J_\alpha &= \frac{\partial}{\partial\alpha'}\; E\left[\varphi_\tau(Y - D'\alpha - X'\beta(\tau))\Psi\right]\Big|_{\alpha=\alpha(\tau)} \\
&= E[f_\epsilon(0|X,Z,D)\Psi D'].
\end{aligned} \tag{7.17}
$$

In other words, for any $\alpha_n \to_p \alpha(\tau)$

$$
\begin{aligned}
\sqrt{n}(\widehat{\vartheta}(\alpha_n,\tau) - \vartheta(\tau)) = &- J_\vartheta^{-1}\mathbb{G}_n\left[f(W,\alpha(\tau),\vartheta(\tau))\right] \\
&- J_\vartheta^{-1}J_\alpha[1 + o_p(1)]\sqrt{n}(\alpha_n - \alpha(\tau)) + o_p(1),
\end{aligned} \tag{7.18}
$$

so

$$
\begin{aligned}
\sqrt{n}(\widehat{\beta}(\alpha_n,\tau) - \beta(\tau)) = &- \bar{J}_\beta\mathbb{G}_n\left[f(W,\alpha(\tau),\vartheta(\tau))\right] \\
&- \bar{J}_\beta J_\alpha[1 + o_p(1)]\sqrt{n}(\alpha_n - \alpha(\tau)) + o_p(1),
\end{aligned} \tag{7.19}
$$

and

$$
\begin{aligned}
\sqrt{n}(\widehat{\gamma}(\alpha_n,\tau) - 0) = &- \bar{J}_\gamma\mathbb{G}_n\left[f(W,\alpha(\tau),\vartheta(\tau))\right] \\
&- \bar{J}_\gamma J_\alpha[1 + o_p(1)]\sqrt{n}(\alpha_n - \alpha(\tau)) + o_p(1),
\end{aligned} \tag{7.20}
$$

where $[\bar{J}_\beta', \bar{J}_\gamma']'$ is the conformable partition of J_ϑ^{-1}.

Center a shrinking closed ball B_n at 0, so that by consistency obtained in Step 2, $\alpha_n - \alpha(\tau) \in B_n$ wp $\to 1$. Then wp $\to 1$

$$
\widehat{\alpha}(\tau) = \underset{\alpha_n - \alpha(\tau)\in B_n}{\operatorname{arginf}}\; W_n(\alpha_n). \tag{7.21}
$$

Note that $\mathbb{G}_n\left[f(W,\alpha(\tau),\vartheta(\tau))\right] \to_d N(0,S)$ by the Central Limit Theorem. Hence $\mathbb{G}_n\left[f(W,\alpha(\tau),\vartheta(\tau))\right] = O_p(1)$, and

$$
\begin{aligned}
W_n(\alpha_n) = &\left[O_p(1) - \bar{J}_\gamma J_\alpha[1 + o_p(1)]\sqrt{n}(\alpha_n - \alpha(\tau))\right]' \\
&\times \left[A(\alpha_n) + o_p(1)\right] \\
&\times \left[O_p(1) - \bar{J}_\gamma J_\alpha[1 + o_p(1)] \times \sqrt{n}(\alpha_n - \alpha(\tau))\right].
\end{aligned} \tag{7.22}
$$

It then follows from (7.21) and (7.22) that $\sqrt{n}(\widehat{\alpha}(\tau) - \alpha(\tau)) = O_p(1)$ since $\bar{J}_\gamma J_\alpha$ has full column rank and $A(\alpha)$ has full rank from R4 and R5. Thus, we have that

$$
\sqrt{n}(\hat{\alpha}(\tau) - \alpha(\tau)) = \arg\inf_{z\in\sqrt{n}B_n}\left[Q_n(z) + o_p(1)\right], \tag{7.23}
$$

where $Q_n(z) := \big(-\bar{J}_\gamma\mathbb{G}_n f(W,\alpha(\tau),\vartheta(\tau)) - \bar{J}_\gamma J_\alpha z\big)' A(\alpha)\big(-\bar{J}_\gamma\mathbb{G}_n f(W,\alpha(\tau),\vartheta(\tau)) - \bar{J}_\gamma J_\alpha z\big)$.

LEMMA 1 (*Approximate Argmins, Knight (1999)*)**.** *Define* Z_n *such that* $Q_n(Z_n) \leq \inf_{z\in\mathbb{R}^d} Q_n(z) + \epsilon_n$, $\epsilon_n \searrow 0$, *and defined* Z_n^* *as* $\arg\inf_{z\in\mathbb{R}^d} Q_n(z)$. *Suppose that* $Z_n = O_p(1)$, $Z_n^* = O_p(1)$, $Z_\infty := \text{argmin}_{z\in\mathbb{R}^d} Q_\infty(z)$ *is uniquely defined in* $\mathbb{R}^d$ *a.s., and* $Q_n(\cdot) \Rightarrow Q_\infty(\cdot)$ *in* $\ell^\infty(K)$ *over any compact sets* K, *where* Q_∞ *is continuous. Then* $Z_n = Z_n^* + o_p(1)$ *and* $Z_n \rightarrow_d Z_\infty$.

Apply Lemma 1 to $Q_n(z)$ defined above and conclude that

$$\sqrt{n}(\widehat{\alpha}(\tau) - \alpha(\tau)) = \arg \inf_{z\in\mathbb{R}^{\dim(\alpha)}} [Q_n(z)] + o_p(1), \tag{7.24}$$

that is

$$\sqrt{n}(\widehat{\alpha}(\tau) - \alpha(\tau)) = -\Big(J_\alpha' \bar{J}_\gamma' A(\alpha(\tau))\bar{J}_\gamma J_\alpha\Big)^{-1}\Big(J_\alpha' \bar{J}_\gamma' A(\alpha(\tau))\bar{J}_\gamma\Big) \times \mathbb{G}_n\left[f(W, \alpha(\tau), \vartheta(\tau), \tau)\right] + o_p(1). \tag{7.25}$$

Hence

$$\sqrt{n}(\widehat{\vartheta}(\widehat{\alpha}(\tau), \tau) - \vartheta(\tau)) = -J_\vartheta^{-1}\Big[I - J_\alpha\Big(J_\alpha' \bar{J}_\gamma' A(\alpha(\tau))\bar{J}_\gamma J_\alpha\Big)^{-1} J_\alpha' \bar{J}_\gamma' A(\alpha(\tau))\bar{J}_\gamma\Big] \times \mathbb{G}_n\left[f(W, \alpha(\tau), \vartheta(\tau), \tau)\right] + o_p(1) \tag{7.26}$$

The conclusion of Proposition 2 follows from $\mathbb{G}_n\left[f(W, \alpha(\tau), \vartheta(\tau))\right] \rightarrow_d N(0, S)$.

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(1), 91–117.

Ambrosetti, A., and G. Prodi (1995): *A primer of nonlinear analysis*, vol. 34 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Amemiya, T. (1977): "The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model," *Econometrica*, 45(4), 955–968.

——— (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711.

Andrews, D. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle, and D. McFadden. North Holland.

Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1993): *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press, Baltimore, MD, Johns Hopkins Series in the Mathematical Sciences.

Chaudhuri, P., K. Doksum, and A. Samarov (1997): "On average derivative quantile regression," *Ann. Statist.*, 25(2), 715–744.

Chen, L., and S. Portnoy (1996): "Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models," *Comm. Statist. Theory Methods*, 25(5), 1005–1032.

Chernozhukov, V., and C. Hansen (2004a): "The Effects of 401(k) Participation on the Wealth Distribution: An Instrumental Quantile Regression Analysis," *Review of Economics and Statistics*, 86(3), 735–751.

——— (2004b): "An IV Model of Quantile Treatment Effects," forthcoming Econometrica.

Chernozhukov, V., and H. Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.

Doksum, K. (1974): "Empirical probability plots and statistical inference for nonlinear models in the two-sample case," *Ann. Statist.*, 2, 267–277.

Graddy, K. (1995): "Testing for Imperfect Competition at the Fulton Fish Market," *Rand Journal of Economics*, 26(1), 75–92.

Gutenbrunner, C., and J. Jurečková (1992): "Regression rank scores and regression quantiles," *Ann. Statist.*, 20(1), 305–330.

Hausman, J. A., and J. G. Sidak (2004): "Why Do the Poor and the Less-Educated Pay Higher Prices for Long-Distance Calls?," *Contributions to Economic Analysis & Policy*, Vol. 3, No. 1, Article 3. http://www.bepress.com/bejeap/contributions/vol3/iss1/art3.

He, X., and L. Zhu (2003): "A lack-of-fit test for quantile regression," *Journal of the American Statistical Association*, to appear.

Januszewski, S. I. (2004): "The Effect of Air Traffic Delays on Airline Prices," UCSD Working Paper. http://weber.ucsd.edu/ sjanusze/www/airtrafficdelays.pdf.

Knight, K. (1998): "Limiting distributions for L_1 regression estimators under general conditions," *Ann. Statist.*, 26(2), 755–770.

——— (1999): "Epi-convergence and Stochastic Equisemicontinuity," Preprint.

Koenker, R. (1994): "Confidence intervals for regression quantiles," in *Asymptotic statistics (Prague, 1993)*, pp. 349–359. Physica, Heidelberg.

Koenker, R., and G. S. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and J. A. F. Machado (1999): "Goodness of fit and related inference processes for quantile regression," *J. Amer. Statist. Assoc.*, 94(448), 1296–1310.

Koenker, R., and S. Portnoy (1987): "L-estimation for linear models," *J. Amer. Statist. Assoc.*, 82(399), 851–857.

Mas-Colell, A. (1979): "Homeomorphisms of compact, convex sets and the Jacobian matrix," *SIAM J. Math. Anal.*, 10(6), 1105–1109.

Portnoy, S. (1991): "Asymptotic behavior of regression quantiles in nonstationary, dependent cases," *J. Multivariate Anal.*, 38(1), 100–113.

Portnoy, S. (2001): "Censored Regression Quantiles," prepirint, www.stat.uiuc.edu.

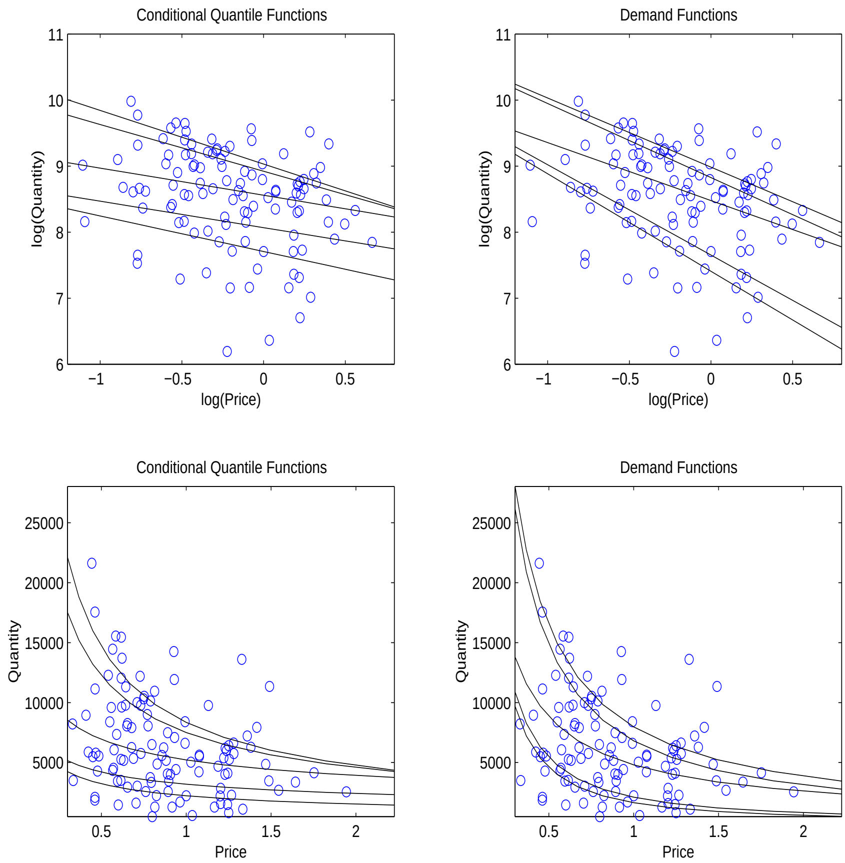
Portnoy, S., and R. Koenker (1997): "The Gaussian Hare and the Laplacian Tortoise," *Statistical Science*, 12, 279–300.

Sakata, S. (2001): "Instrumental Variable Estimation Based on the Least Absolute Deviation Estimator," Preprint, Department of Economics, University of Michigan.

van der Vaart, A. W. (1998): *Asymptotic statistics*. Cambridge University Press, Cambridge.

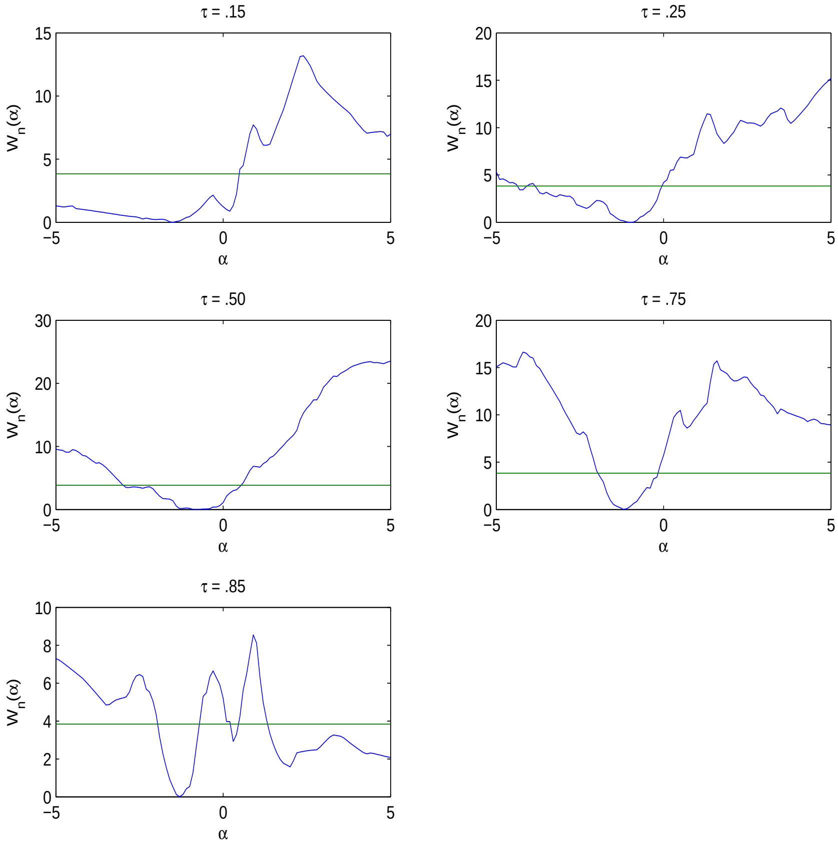
van der Vaart, A. W., and J. A. Wellner (1996): *Weak convergence and empirical processes*. Springer-Verlag, New York.

Figure 1. Estimates of Effect of Price on Quantity by QR and IVQR



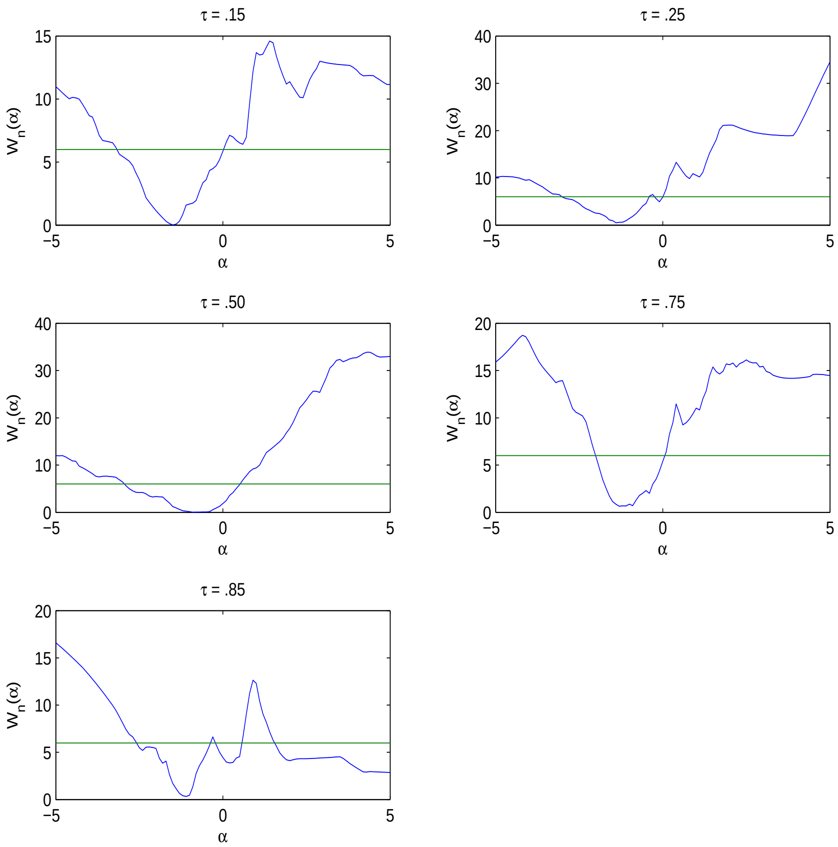
Note: *Left Column:* The estimated conditional quantile curves of the quantity of fish sold as a function of price for $\tau = .15, .25, .50, .75$, and $.85$. The top display is in log-price log-quantity space with log-price on the horizontal axis and log-quantity on the vertical axis. The bottom display is in price-quantity space with price on the horizontal axis and quantity on the vertical axis. *Right Column:* The demand curves estimated by IVQR for $\tau = .15, .25, .50, .75$, and $.85$. The top display is in log-price log-quantity space with log-price on the horizontal axis and log-quantity on the vertical axis. The bottom display is in price-quantity space with price on the horizontal axis and quantity on the vertical axis.

Figure 2. Statistic $W_n(\alpha)$ in Demand Example Using Stormy as an Instrument



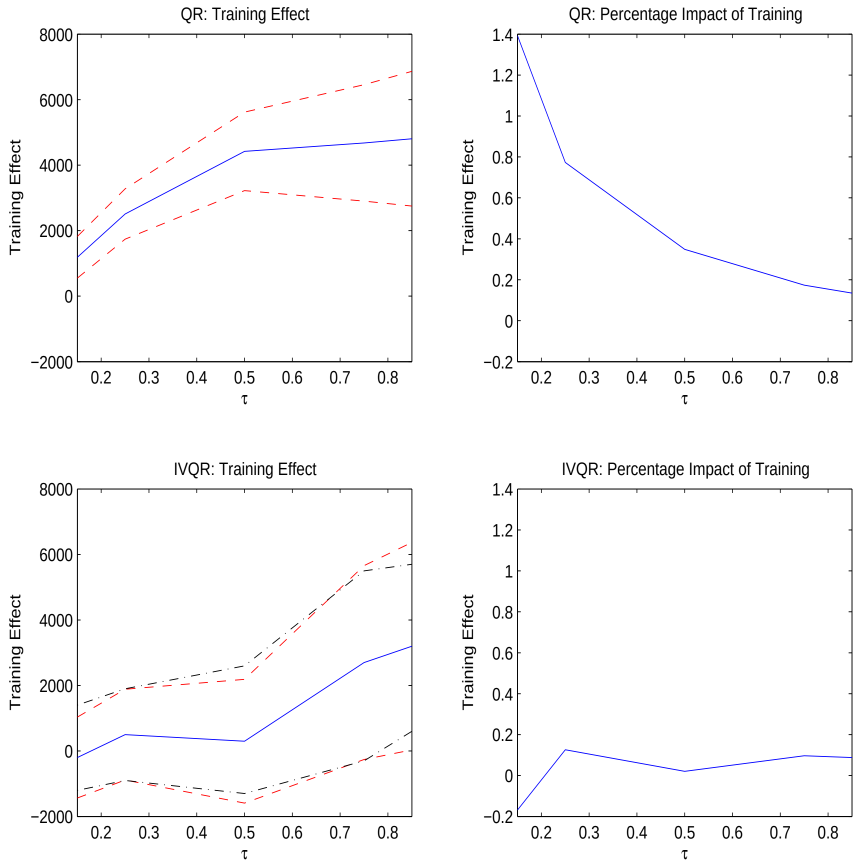
Note: Objective functions and dual confidence regions for demand for fish example. All models are as specified in the main text. The estimates make use of one instrument, *Stormy*. α is on the horizontal axis and $W_n(\alpha)$ is on the vertical axis. The horizontal line is the 95% critical value from a χ^2_1. The dual confidence region is all values of α such that the $W_n(\alpha)$ lies below the horizontal line.

Figure 3. Statistic $W_n(\alpha)$ in Demand Example Using Stormy and Mixed as Instruments



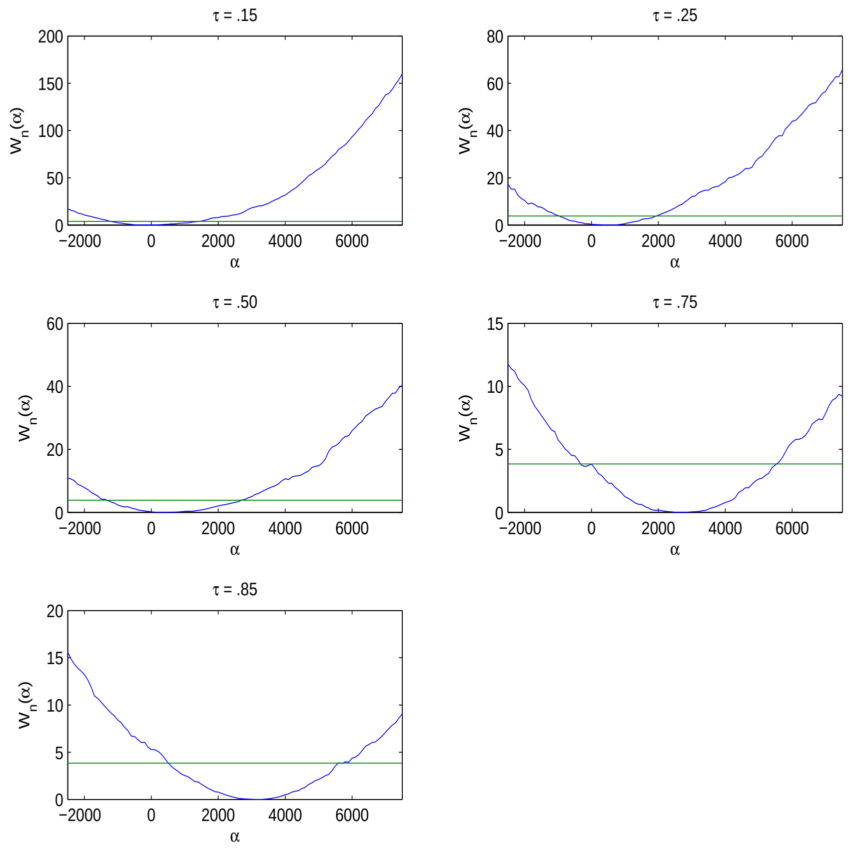
Note: Objective functions and dual confidence regions for demand for fish example. All models are as specified in the main text. The estimates make use of two instruments, *Stormy* and *Mixed*. α is on the horizontal axis and $W_n(\alpha)$ is on the vertical axis. The horizontal line is the 95% critical value from a χ^2_2. The dual confidence region is all values of α such that $W_n(\alpha)$ lies below the horizontal line.

Figure 4. Estimates of the Training Impact by QR and by IVQR



Note: *Left Column:* QR and IVQR estimates of the impact of a job training program on earnings for $\tau = .15$, .25, .50, .75, and .85. The top panel reports the QR estimate of the training impact, and the bottom panel reports the IVQR results. In each figure, the solid line represents the point estimates, and the dashed (- -) line represents the 95% confidence interval formed using the direct inference approach. For the IVQR results, the dash-dot (-.) line represents the 95% confidence bound constructed using the dual inference procedure described in the text. In both figures, the horizontal axis measures the quantile index τ, and the vertical axis is the impact of training on earning quantiles measured in dollars. Models include covariates as specified in the text, and the sample size is 5,102. *Right Column:* QR and IVQR estimates of the percentage impact of training for $\tau = .15$, .25, .50, .75, and .85. The top panel reports the QR estimate of the training impact, and the bottom panel reports the IVQR results. Percentage impacts are for moving from non-training to training and all other covariates are evaluated at their sample mean. In both figures, the horizontal axis measures the quantile index τ, and the vertical axis is the percentage impact of training.

Figure 5. Statistic $W_n(\alpha)$ in the Training Example.



Note: Objective functions and dual confidence regions for returns to training example. All models are as specified in the main text. The estimates use random assignment to the training program as the instrument. α is on the horizontal axis and $W_n(\alpha)$ is on the vertical axis. The horizontal line is the 95% critical value from a χ^2_1. The dual confidence region is all values of α such that the function value lies below the horizontal line.