FINITE SAMPLE INFERENCE FOR
QUANTILE REGRESSION MODELS

Victor Chernozhukov
Christian Hansen
Michael Jansson

FINITE SAMPLE INFERENCE FOR QUANTILE REGRESSION MODELS

VICTOR CHERNOZHUKOV   CHRISTIAN HANSEN   MICHAEL JANSSON

ABSTRACT. Under minimal assumptions finite sample confidence bands for quantile regression models can be constructed. These confidence bands are based on the "conditional pivotal property" of estimating equations that quantile regression methods aim to solve and will provide valid finite sample inference for both linear and nonlinear quantile models regardless of whether the covariates are endogenous or exogenous. The confidence regions can be computed using MCMC, and confidence bounds for single parameters of interest can be computed through a simple combination of optimization and search algorithms. We illustrate the finite sample procedure through a brief simulation study and two empirical examples: estimating a heterogeneous demand elasticity and estimating heterogeneous returns to schooling. In all cases, we find pronounced differences between confidence regions formed using the usual asymptotics and confidence regions formed using the finite sample procedure in cases where the usual asymptotics are suspect, such as inference about tail quantiles or inference when identification is partial or weak. The evidence strongly suggests that the finite sample methods may usefully complement existing inference methods for quantile regression when the standard assumptions fail or are suspect.

Key Words: Quantile Regression, Extremal Quantile Regression, Instrumental Quantile Regression

1. INTRODUCTION

Quantile regression (QR) is a useful tool for examining the effects of covariates on an outcome variable of interest; see e.g. Koenker (2005). Perhaps the most appealing feature of QR is that it allows estimation of the effect of covariates on many points of the outcome distribution

including the tails as well as the center of the distribution. While the central effects are useful summary statistics of the impact of a covariate, they do not capture the full distributional impact of a variable unless the variable affects all quantiles of the outcome distribution in the same way. Due to its ability to capture heterogeneous effects and its interesting theoretical properties, QR has been used in many empirical studies and has been studied extensively in theoretical econometrics; see especially Koenker and Bassett (1978), Portnoy (1991), Buchinsky (1994), and Chamberlain (1994), among others.

In this paper, we contribute to the existing literature by considering finite sample inference for quantile regression models. We show that valid finite sample confidence regions can be constructed for parameters of a model defined by quantile restrictions under minimal assumptions. These assumptions do not require the imposition of distributional assumptions and will be valid for both linear and nonlinear conditional quantile models and for models which include endogenous as well as exogenous variables. The basic idea of the approach is to make use of the fact that the estimating equations that correspond to conditional quantile restrictions are conditionally pivotal; that is, conditional on the exogenous regressors and instruments, the estimating equations are pivotal in finite samples. Thus, valid finite sample tests and confidence regions can be constructed based on these estimating equations.

The approach we pursue is related to early work on finite sample inference for the sample (unconditional) quantiles. The existence of finite sample pivots is immediate for unconditional quantiles as illustrated, for example, in Walsh (1960) and MacKinnon (1964). However, the existence of such pivots in the regression case is less obvious. We extend the results from the unconditional case to the estimation of conditional quantiles by noting that conditional on the exogenous variables and instruments the estimating equations solved by QR methods are pivotal in finite samples. This property suggests that tests based on these quantities can be used to obtain valid finite sample inference statements. The resulting approach is similar in spirit to the rank-score methods, e.g. Gutenbrunner, Jurečková, Koenker, and Portnoy (1993), but does not require asymptotics or homoscedasticity for its validity.

The finite sample approach that we develop has a number of appealing features. The approach will provide valid inference statements under minimal assumptions, essentially requiring some weak independence assumptions on sampling mechanisms and continuity of conditional quantile functions in the probability index. In endogenous settings, the finite sample approach will remain valid in cases of weak identification or partial identification (e.g. as in Tamer (2003)). In this sense, the finite sample approach usefully complements asymptotic approximations and can be used in situations where the validity of the assumptions necessary to justify these approximations is questionable.

The chief difficulty with the finite sample approach is computational. In general, implementing the approach will require inversion of an objective function-like quantity which may be quite difficult if the number of parameters is large. To help alleviate this computational problem, we explore the use of Markov Chain Monte Carlo (MCMC) methods for constructing joint confidence regions. The use of MCMC allows us to draw an adaptive set of grid points which offers potential computational gains relative to more naive grid based methods. We also consider a simple combination of search and optimization routines for constructing marginal confidence bounds. When interest focuses on a single parameter, this approach may be computationally convenient and may be more robust in nonregular situations than an approach aimed at constructing the joint confidence region.

Another potential disadvantage of the proposed finite sample approach is that one might expect that minimal assumptions would lead to wide confidence intervals. We show that this concern is unwarranted for joint inference: The finite sample tests have correct size and good asymptotic power properties. However, conservativity may be induced by going from joint to marginal inference by projection methods. In this case, the finite sample confidence bounds may not be sharp.

To explore these issues, we examine the properties of the finite sample approach in simulation and empirical examples. In the simulations, we find that joint tests based on the finite sample procedure have correct size while conventional asymptotic tests tend to be size distorted and are severely size distorted in some cases. We also find that finite sample tests about individual regression parameters have size less than the nominal value, though they have reasonable power in many situations. On the other hand, the asymptotic tests tend to have size that is greater than the nominal level.

We also consider the use of finite sample inference in two empirical examples. In the first, we consider estimation of a demand curve in a small sample; and in the second, we estimate the returns to schooling in a large sample. In the demand example, we find modest differences between the finite sample and asymptotic intervals when we estimate conditional quantiles not instrumenting for price and large differences when we instrument for price. In the schooling example, the finite sample and asymptotic intervals are almost identical in models in which we treat schooling as exogenous, and there are large differences when we instrument for schooling. These results suggest that the identification of the structural parameters in the instrumental variables models in both cases is weak.

The remainder of this paper is organized as follows. In the next section, we formally introduce the modelling framework we are considering and the basic finite sample inference results. Section 3 presents results from the simulation and empirical examples, and Section 4 concludes.

Asymptotic properties of the finite sample procedure that include asymptotic optimality results are contained in an appendix.

2. FINITE SAMPLE INFERENCE

2.1. **The Model.** In this paper, we consider finite sample inference in the quantile regression model characterized below.

Assumption 1. Let there be a probability space $(\Omega, \mathcal{F}, P)$ and a random vector (Y, D', Z', U) defined on this space, with $Y \in \mathbb{R}$, $D \in \mathbb{R}^{\dim(D)}$, $Z \in \mathbb{R}^{\dim(Z)}$, and $U \in (0, 1)$ P-a.s, such that

    A1 $Y = q(D, U)$ for a function $q(d, u)$ that is measurable.
    A2 $q(d, u)$ is strictly increasing in u for each d in the support of D.
    A3 $U \sim \text{Uniform}(0, 1)$ and is independent from Z.
    A4 D is statistically dependent on Z.

When $D = Z$, the model in A1-4 corresponds to the conventional quantile regression model with exogenous covariates, cf. Koenker (2005) where Y is the dependent variable, D is the regressor, and $q(d, \tau)$ is the τ-quantile of Y conditional on $D = d$ for any $\tau \in (0, 1)$. In this case, A1, A3, and A4 are not restrictive and provide a representation of Y, while A2 restricts Y to have a continuous distribution function. The exogenous model was introduced in Doksum (1974) and Koenker and Bassett (1978). It usefully generalizes the classical linear model $Y = D'\gamma_0 + \gamma_1(U)$ by allowing for quantile specific effects of covariates D. Estimation and asymptotic inference for the linear version of this model, $Y = D'\theta(U)$, was developed in Koenker and Bassett (1978), and estimation and inference results have been extended in a number of useful directions by subsequent. Matzkin (2003) provides many economic examples that fall in this framework and considers general nonparametric methods for asymptotic inference.

When $D \neq Z$ but Z is a set of instruments that are independent of the structural disturbance U, the model A1-4 provides a generalization of the conventional quantile model that allows for endogeneity. See Chernozhukov and Hansen (2001, 2005a, 2005b) for discussion of the model as well as for semi-parametric estimation and inference theory under strong and weak identification. See Chernozhukov, Imbens, and Newey (2006) for a nonparametric analysis of this model and Chesher (2003) for a related nonseparable model. The model A1-4 can be thought of as a general nonseparable structural model that allows for endogenous variables as well as a treatment effects model with heterogeneous treatment effects. In this case, D and U may be jointly determined rendering the conventional quantile regression invalid for making inference on the structural quantile function $q(d, \tau)$. This model 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.

Note that in this case, A4 is necessary for identification. However, the finite sample inference results presented below will remain valid even when A4 is not satisfied.

Under Assumption 1, we state the following result which will provide the basis for the finite sample inference results that follow.

Proposition 1 (Main Statistical Implication). *Suppose A1-3 hold, then*

$$
\textbf{\textit{1.}} \quad P[Y \leq q(D, \tau)|Z] = \tau, \tag{2.1}
$$

$$
\textbf{\textit{2.}} \quad \{Y \leq q(D, \tau)\} \text{ is Bernoulli}(\tau) \text{ conditional on } Z . \tag{2.2}
$$

PROOF: $\{Y \leq q(D, \tau)\}$ *is equivalent to* $\{U \leq \tau\}$ *which is independent of* Z. *The results then follow from* $U \sim U(0, 1)$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Equation (2.1) provides a set of moment conditions that can be used to identify and estimate the quantile function $q(d, \tau)$. When $D = Z$, these are the standard moment conditions used in quantile regression which have been analyzed extensively starting with Koenker and Bassett (1978) and when $D \neq Z$, the identification and estimation of $q(d, \tau)$ from (2.1) is considered in Chernozhukov and Hansen (2005b).

(2.2) is the key result from which we obtain the finite sample inference results. The result states that the event $\{Y \leq q(D, \tau)\}$ conditional on Z is distributed exactly as a Bernoulli(τ) random variable regardless of the sample size. This random variable depends only on τ which is known and so is pivotal in finite samples. These results allow the construction of exact finite sample confidence regions and tests conditional on the observed data, Z.

2.2. **Model and Sampling Assumptions.** In the preceding section, we outlined a general heterogeneous effect model and discussed how the model relates to quantile regression. We also showed that the model implies that $\{Y \leq q(D, \tau)\}$ conditional on Z is distributed exactly as a Bernoulli(τ) random variable in finite samples. In order to operationalize the finite sample inference, we also impose the following conditions.

Assumption 2. Let $\tau \in (0, 1)$ denote the quantile of interest. Suppose that there is a sample $(Y_i, D_i, Z_i, i = 1, ..., n)$ on probability space $(\Omega, \mathcal{F}, P)$ (possibly dependent on the sample size), such that A1-A4 holds for each $i = 1, ..., n$, and the following additional conditions hold:

- **A5 (Finite-Parameter Model):** $q(D, \tau) = q(D, \theta_0, \tau)$, for some $\theta_0 \in \Theta_n \subset \mathbb{R}^{K_n}$, where the function $q(D, \theta, \tau)$ is known, but θ_0 is not.
- **A6 (Conditionally Independent Sampling):** $(U_1, ..., U_n)$ are i.i.d. Uniform(0, 1), conditional on $(Z_1, ..., Z_n)$.

We will use the letter $\mathcal{P}$ to denote all probability laws P on the measure space $(\Omega, \mathcal{F})$ that satisfy conditions A1-6.

Conditions A5-6 restrict the model A1-4 sufficiently to allow finite sample inference. A5 requires that the τ-quantile function $q(d, \tau)$ is known up to a finite-dimensional parameter θ_0 (where θ_0 may vary with τ). Since we are interested in finite sample inference, it is obvious that such a condition is needed. However, A5 does allow for the model to depend on the sample size n in the Pitman sense, and allows the dimension of the model, K_n, to increase with n in the sense of Huber (1973) and Portnoy (1985) where $K_n \to \infty$ as $n \to \infty$. In this sense, we can allow flexible (approximating) functional forms for $q(D, \theta_0, \tau)$ such as linear combinations of B-splines, trigonometric, power, and spline series. Condition A6 is satisfied if $(Y_i, X_i, Z_i, i = 1, ..., n)$ are i.i.d., but more generally allows rather rich dynamics, e.g. of the kinds considered in Koenker and Xiao (2004a) and Koenker and Xiao (2004b).

2.3. The Finite Sample Inference Procedure.

Using the conditions discussed in the previous sections, we are able to provide the key results on finite sample inference. We start by noting that equation (2.1) in Proposition 1 justifies the following generalized method-of-moments (GMM) function for estimating θ_0:

$$ L_n(\theta) = \frac{1}{2} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_i(\theta) \right]' W_n \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m_i(\theta) \right], \tag{2.3} $$

where $m_i(\theta) = [\tau - 1(Y_i \leq q(D_i, \theta, \tau))] \, g(Z_i)$. In this expression, $g(Z_i)$ is a known vector of functions of Z that satisfies $\dim(g(Z)) \geq \dim(\theta_0)$, and W_n is a positive definite weight matrix, which is fixed conditional on $Z_1, ..., Z_n$. A convenient and natural choice of W_n is given by

$$ W_n = \frac{1}{\tau(1-\tau)} \left[\frac{1}{n} \sum_{i=1}^{n} g(Z_i)g(Z_i)' \right]^{-1}, $$

which equals the inverse of the variance of $n^{-1/2} \sum_{i=1}^{n} m_i(\theta_0)$ conditional on $Z_1, ..., Z_n$. Since this conditional variance does not depend on θ_0, the GMM function with W_n defined above also corresponds to the continuous-updating estimator of Hansen, Heaton, and Yaron (1996).

We focus on the GMM function $L_n(\theta)$ for defining the key results for finite sample inference. The GMM function provides an intuitive statistic for performing inference given its close relation to standard estimation and asymptotic inference procedures. In addition, we show in the appendix that testing based on $L_n(\theta)$ may have useful asymptotic optimality properties.

We now state the key finite sample results.

Proposition 2. *Under A1-A6, statistic $L_n(\theta_0)$ is conditionally pivotal: $L_n(\theta_0) \overset{d}{=} \mathcal{L}_n$, conditional on $(Z_1, ..., Z_n)$, where*

$$\mathcal{L}_n = \frac{1}{2} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tau - B_i) \cdot g(Z_i) \right)' W_n \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tau - B_i) \cdot g(Z_i) \right),$$

and $(B_1, ..., B_n)$ are iid Bernoulli rv's with $EB_i = \tau$, which are independent of $(Z_1, ..., Z_n)$.

PROOF: *Implication 2 of Proposition 1 and A6 imply the result.* $\qquad\square$

Proposition 2 formally states the finite sample distribution of the GMM function $L_n(\theta)$ at $\theta = \theta_0$. Conditional on $(Z_1, ..., Z_n)$, the distribution does not depend on any unknown parameters, and appropriate critical values from the distribution may be obtained allowing finite sample inference on θ_0.

Given the finite sample distribution of $L_n(\theta_0)$, a $1 - \alpha$-level test of the null hypothesis that $\theta = \theta_0$ is given by the rule that rejects the null if $L_n(\theta) > c_n(\alpha)$ where $c_n(\alpha)$ is the α-quantile of $\mathcal{L}_n$. By inverting this test-statistic, one obtains confidence regions for θ_0.

Let $CR(\alpha)$ be the $c_n(\alpha)$-level set of the function $L_n(\theta)$: $CR(\alpha) \equiv \{\theta : L_n(\theta) \le c_n(\alpha)\}$. It follows immediately from the previous results that $CR(\alpha)$ is a valid α-level confidence region for θ_0. This result is stated formally in Proposition 3.

Proposition 3. *Fix an $\alpha \in (0, 1)$. $CR(\alpha)$ is a valid α-level confidence region for inference about θ_0 in finite samples: $\Pr_P(\theta_0 \in CR(\alpha)) \ge \alpha$. $CR(\alpha)$ is also a valid critical region for obtaining a $1 - \alpha$-level test of $\theta = \theta_0$: $\Pr_P(\theta_0 \notin CR(\alpha)) \le 1 - \alpha$. Moreover, these results hold uniformly in $P \in \mathcal{P}$, $\inf_{P \in \mathcal{P}} \Pr_P(\theta_0 \in CR(\alpha)) \ge \alpha$ and $\sup_{P \in \mathcal{P}} \Pr_P(\theta_0 \notin CR(\alpha)) \le 1 - \alpha$.*

PROOF: *$\theta_0 \in CR(\alpha)$ is equivalent to $\{L_n(\theta_0) \le c_n(\alpha)\}$ and $\Pr_P\{L_n(\theta_0) \le c_n(\alpha)\} \ge \alpha$, by the definition of $c_n(\alpha) := \inf\{l : P\{\mathcal{L}_n \le l\} \ge \alpha\}$ and $L_n(\theta_0) =_d \mathcal{L}_n$.* $\qquad\square$

Proposition 3 demonstrates how one may obtain valid finite sample confidence regions and tests for the parameter vector θ characterizing the quantile function $q(D, \theta_0, \tau)$. Thus, this result generalizes the approach of Walsh (1960) from the sample quantiles to the regression case. It is also apparent that the pivotal nature of the finite sample approach is similar to the asymptotically pivotal nature of the rank-score method, cf. Gutenbrunner, Jurečková, Koenker, and Portnoy (1993) and Koenker (1997), and the bootstrap method of Parzen, Wei, and Ying (1994).[1] In sharp contrast to these methods, the finite sample approach does not

---

[1]The finite sample method should not be confused with the Gibbs bootstrap proposed in He and Hu (2002) who propose a computationally attractive variation on Parzen, Wei, and Ying (1994). The method is also very different from specifying the finite sample density of quantile regression as in Koenker and Bassett (1978). The

rely on asymptotics and is valid in finite samples. Moreover, the rank-score method relies on a homoscedasticity assumption, while the finite sample approach does not.

The statement of Proposition 3 is for joint inference about the entire parameter vector. One can define a confidence region for a real-valued functional $\psi(\theta_0, \tau)$ as

$$CR(\alpha, \psi) = \{\psi(\theta, \tau) : \theta \in CR(\alpha)\}.$$

Since the event $\{\theta_0 \in CR(\alpha)\}$ implies the event $\{\psi(\theta_0, \tau) \in CR(\alpha, \psi)\}$, it follows that $\inf_{P \in \mathcal{P}} \Pr_P(\psi(\theta_0, \tau) \in CR(\alpha, \psi)) \geq \alpha$ by Proposition 3. For example, if one is interested in inference about a single component of θ, say $\theta_{[1]}$, a confidence region for $\theta_{[1]}$ may be constructed as the set $\{\theta_{[1]} : \theta \in CR(\alpha)\}$. That is, the confidence region for $\theta_{[1]}$ is obtained by first taking all vectors of θ in $CR(\alpha)$ and then extracting the element from each vector corresponding to $\theta_{[1]}$. Confidence bounds for $\theta_{[1]}$ may be obtained by taking the infimum and supremum over this set of values for $\theta_{[1]}$.

2.4. **Primary Properties of the Finite Sample Inference.** The finite sample tests and confidence regions obtained in the preceding section have a number of interesting and appealing features. Perhaps the most important feature of the proposed approach is that it allows for finite sample inference under weak conditions. Working with a model defined by quantile restrictions makes it possible to construct exact joint inference in a general non-linear, non-separable model with heterogeneous effects that allows for endogeneity. This is in contrast with many other inference approaches for instrumental variables models that are developed for additive models only.

The approach is valid without imposing distributional assumptions and allows for general forms of heteroskedasticity and rich forms of dynamics. The result is obtained without relying on asymptotic arguments and essentially requires only that Y has a continuous conditional distribution function given Z. In contrast with conventional asymptotic approaches to inference in quantile models, the validity of the finite sample approach does not depend upon having a well-behaved density for Y: It does not rely on the density of Y given $D = d$ and $Z = z$ being continuous or differentiable in y or having connected support around $q(d, \tau)$, as required e.g. in Chernozhukov and Hansen (2001).

In addition to these features, the finite sample inference procedure will remain valid in situations where the parameters of the model are only partially identified. The confidence regions obtained from the finite sample procedure will provide valid inference about $q(D, \tau) = q(D, \theta_0, \tau)$ even when θ_0 is not uniquely identified by $P[Y \leq q(D, \theta_0, \tau)|Z] = \tau$. This builds on the point made in Hu (2002). In addition, since the inference is valid for any n, it follows

---

finite sample density of QR is not pivotal and can not be used for finite sample inference unless the nuisance parameters (the conditional density of the residuals given the regressors) are specified.

trivially that it remains valid under the asymptotic formalization of "weak instruments", as defined e.g. in Stock and Wright (2000).

As noted previously, inference statements obtained from the finite sample procedure will also remain valid in models where the dimension of the parameter space K_n is allowed to increase with future increases of n since the statements are valid for any given n. Thus, the results of the previous section remain valid in the asymptotics of Huber (1973) and Portnoy (1985) where $K_n/n \to 0, K_n \to \infty, n \to \infty$. These rate conditions are considerably weaker than those required for conventional inference using Wald statistics as described in Portnoy (1985) and Newey (1997) which require $K_n^2/n \to 0, K_n \to \infty, n \to \infty$.

Inference statements obtained from the finite sample procedure will be valid for inference about extremal quantiles where the usual asymptotic approximation may perform quite poorly.[2] One alternative to using the conventional asymptotic approximation for extremal quantiles is to pursue an approach explicitly aimed at performing inference for extremal quantiles, e.g as in Chernozhukov (2000). The extreme value approach improves upon the usual asymptotic approximation but requires a regular variation assumption on the tails of the conditional distribution of $Y|D$, that the tail index does not vary with D, and also relies heavily on linearity and exogeneity. None of these assumptions are required in the finite sample approach, so the inference statements apply more generally than those obtained from the extreme value approach.

It is also worth noting that while the approach presented above is explicitly finite sample, it will remain valid asymptotically. Under conventional assumptions and asymptotics, e.g. Pakes and Pollard (1989) and Abadie (1995), the inference approaches conventional GMM based joint inference.

Finally, it is important to note that inference is simultaneous on all components of θ and that for joint inference the approach is not conservative. Inference about subcomponents of θ may be made by projections, as illustrated in the previous section, and may be conservative. We explore the degree of conservativity induced by marginalization in a simulation example in Section 3.

2.5. **Computation.** The main difficulty with the approach introduced in the previous sections is computing the confidence regions. The distribution of $L_n(\theta_0)$ is not standard, but its critical values can be easily constructed by simulation. The more serious problem is that inverting the function $L_n(\theta)$ to find joint confidence regions may pose a significant computational challenge.

---

[2]Whether a given quantile is extremal depends on the sample size and underlying data generating process. However, Chernozhukov (2000) finds that the usual asymptotic approximation behaves quite poorly in some examples for $0 < \tau < .2$ and $1 > \tau > .8$.

One possible approach is to simply use a naive grid-search, but as the dimension of θ increases, this approach becomes intractable. To help alleviate this problem, we explore the use of MCMC methods. MCMC seems attractive in this setting because it generates an adaptive set of grid points and so should explore the relevant region of the parameter space more quickly than performing a conventional grid search. We also consider a marginalization approach that combines a one-dimensional grid search with optimization for estimating a confidence bound for a single parameter which may be computationally convenient and may be more robust than MCMC in some irregular cases.

2.5.1. *Computation of the Critical Value.* The computation of the critical value $c_n(\alpha)$ may proceed in a straightforward fashion by simulating the distribution $\mathcal{L}_n$. We briefly outline a simulation routine below.

Algorithm 1 (Computation of $c_n(\alpha)$.). *Given* $(Z_i, i = 1, ..., n)$, *for* $j = 1, ..., J$: *1. Draw* $(U_{i,j}, i \leq n)$ *as iid Uniform, and let* $(B_{i,j} = 1(U_{i,j} \leq \tau), i \leq n)$. *2. Compute* $\mathcal{L}_{n,j} = \frac{1}{2}(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tau - B_{i,j}) \cdot g(Z_i))' W_n (\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tau - B_{i,j}) \cdot g(Z_i))$. *3. Obtain* $c_n(\alpha)$ *as the α-quantile of the sample* $(\mathcal{L}_{n,j}, j = 1, ..., J)$, *for a large number* J.

2.5.2. *Computation of Confidence Regions.* Finding the confidence region requires computing the $c_n(\alpha)$-level set of the function $L_n(\theta)$ which involves inverting a non-smooth, non-convex function. For even moderate sized problems, the use of a conventional grid search is impractical due to the computational curse of dimensionality.

To help resolve this problem, we consider the use of a generic random walk Metropolis-Hastings MCMC algorithm.[3] The idea is that the MCMC algorithm will generate a set of adaptive grid-points that are placed in relevant regions of the parameter space only. By focusing more on relevant regions of the parameter space, the use of MCMC may alleviate the computational problems associated with a conventional grid search.

To implement the MCMC algorithm, we treat $f(\theta) \propto \exp(-L_n(\theta))$ as a quasi-posterior density and feed it into a random walk MCMC algorithm. (The idea is similar to that in Chernozhukov and Hong (2003), except that we use it here to get level sets of the objective function rather than pseudo-posterior means and quantiles.) The basic random walk MCMC is implemented as follows:

Algorithm 2 (Algorithm 2. Random Walk MCMC.). *For a symmetric proposal density $h(\cdot)$ and given $\theta^{(t)}$, 1. Generate $\theta_{prop}^{(t)} \sim h(\theta - \theta^{(t)})$. 2. Take $\theta^{(t+1)} = \theta_{prop}^{(t)}$ with probability $\min\{1, f(\theta_{prop}^{(t)})/f(\theta^{(t)})\}$ and $\theta^{(t)}$ otherwise. 3. Store $(\theta^{(t)}, \theta_{prop}^{(t)}, L_n(\theta^{(t)}), L_n(\theta_{prop}^{(t)}))$. 4. Repeat Steps 1-3 J times replacing $\theta^{(t)}$ with $\theta^{(t+1)}$ as starting point for each repetition.*

---

[3]Other MCMC algorithms or stochastic search methods could also be employed.

At each step, the MCMC algorithm considers two potential values for θ and obtains the corresponding values of the objective function. Step 3 above differs from a conventional random walk MCMC in that we are interested in every value considered not just those accepted by the procedure.

The implementation of the MCMC algorithm requires the user to specify a starting value for the chain and a transition density $g(\cdot)$. The choice of both quantities can have important practical implications, and implementation in any given example will typically involve some fine tuning in both the choice of $g(\cdot)$ and the starting value.[4] Robert and Casella (1998) provide an excellent overview of these and related issues.

As illustrated above, the MCMC algorithm generates a set of grid points $\{\theta^{(1)}, ..., \theta^{(k)}\}$ and, as a by-product, a set of values for the objective function $\{L_n(\theta^{(1)}), ..., L_n(\theta^{(k)})\}$. Using this set of evaluations of the objective function, we can construct an estimate of the critical region by taking the set of draws for θ where the value of $L_n(\theta) \leq c_n(\alpha)$: $\widetilde{CR}(\alpha) = \{\theta^{(i)} : L_n(\theta^{(i)}) \leq c(\alpha)\}$.

Figure 1 illustrates the construction of these confidence regions. Both figures illustrate 95% confidence regions for $\tau = .5$ in a simple demand example that we discuss in Section 3. The regions illustrated here are for a model in which price is treated as exogenous. Values of the intercept are on the x-axis and values of the coefficient on price are on the y-axis.

Figure 1 Panel A illustrates a set of MCMC draws in this example. Each symbol + represents an MCMC draw of the parameter vector that satisfies $L_n(\theta) \leq c_n(.95)$, and each symbol $\bullet$ represents a draw that does not satisfy this condition. Thus, (a numerical approximation to) the confidence region is given by the area covered with symbol + in the figure. In this case, the MCMC algorithm appears to be doing what we would want. The majority of the draws come from within the confidence region, but the algorithm does appear to do a good job of exploring areas outside of the confidence region as well.

Panel B of Figure 1 presents a comparison of the confidence region obtained through MCMC and the confidence region obtained through a grid search. The boundary of the grid search region is represented by the black line, and the MCMC region is again given by the light gray area in the figure. Here, we see that the two regions are almost identical. Both regions include some points that are not in the other, but the agreement is quite impressive.

2.5.3. *Computation of Confidence Bounds for Individual Regression Parameters.* The MCMC approach outlined above may be used to estimate joint confidence regions which can be used

---

[4]In our applications, we use estimates of θ and the corresponding asymptotic distribution obtained from the quantile regression of Koenker and Bassett (1978) in exogenous cases and from the inverse quantile regression of Chernozhukov and Hansen (2001) in endogenous cases as starting values and transition densities.

for joint inference about the entire parameter vector or for inference about subsets of regression parameters. If one is interested solely in inference about an individual regression parameter, there may be a computationally more convenient approach. In particular, for constructing a confidence bound for a single parameter, knowledge of the entire joint confidence region is unnecessary which suggests that we may collapse the d-dimensional search to a one-dimensional search.

For concreteness, suppose we are interested in constructing a confidence bound for a particular element of θ, denoted $\theta_{[1]}$, and let $\theta_{[-1]}$ denote the remaining elements of the parameter vector. We note that a value of $\theta_{[1]}$, say $\theta_{[1]}^*$, will lie inside the confidence bound as long as there exists a value of θ with $\theta_{[1]} = \theta_{[1]}^*$ that satisfies $L_n(\theta) \leq c_n(\alpha)$. Since only one such value of θ is required to place $\theta_{[1]}^*$ in the confidence bound, we may restrict consideration to θ^*, the point that minimizes $L_n(\theta)$ conditional on $\theta_{[1]} = \theta_{[1]}^*$. If $L_n(\theta^*) > c_n(\alpha)$, we may conclude that there will be no other point that satisfies $L_n(\theta) \leq c_n(\alpha)$ and exclude $\theta_{[1]}^*$, from the confidence bound. On the other hand, if $L_n(\theta^*) \leq c_n(\alpha)$, we have found a point that satisfies $L_n(\theta) \leq c_n(\alpha)$ and can include $\theta_{[1]}^*$ in the confidence bound.

This suggests that a confidence bound for $\theta_{[1]}$ can be constructed using the following simple algorithm that combines a one-dimensional grid search with optimization.

Algorithm 3 (Marginal Approach.). *1. Define a suitable set of values for $\theta_{[1]}$, $\{\theta_{[1]}^j$, $j = 1, ..., J\}$. 2. For $j = 1, ..., J$, find $\theta_{[-1]}^j = \arg \inf_{\theta_{[-1]}} L_n((\theta_{[1]}^j, \theta_{[-1]}')')$. 3. Calculate the confidence region for $\theta_{[1]}$ as $\{\theta_{[1]}^j : L_n((\theta_{[1]}^j, \theta_{[-1]}^{j\prime})') \leq c_n(\alpha)\}\}$.*

In addition to being computationally convenient for finding confidence bounds for individual parameters in high-dimensional settings, we also anticipate that this approach will perform well in some irregular cases. Since the marginal approach focuses on only one parameter, it will typically be easy to generate a tractable and reasonable search region. The approach will have some robustness to multimodal objective functions and potentially disconnected confidence sets because it considers all values in the grid search region and will not be susceptible to getting stuck at a local mode.

3. Simulation and Empirical Examples

In the preceding section, we presented an inference procedure for quantile regression that provides exact finite sample inference for joint hypotheses and discussed how confidence bounds for subsets of quantile regression parameters may be obtained. In the following, we further explore the properties of the proposed finite sample approach through brief simulation examples

and through two simple case studies.[5] In the simulations, we find that tests about the entire parameter vector based on the finite sample method have the correct size while tests which make use of the asymptotic approximation may be substantially size distorted. When we consider marginal inference, we find that using the asymptotic approximation leads to tests that reject too often while, as would be expected, the finite sample method yields conservative tests.

We also consider the use of the finite sample inference procedure in two case studies. In the first, we consider estimation of a demand model in a small sample; and in the second, we consider estimation of the impact of schooling on wages in a rather large sample. In both cases, we find that the finite sample and asymptotic intervals are similar when the variables of interest, price and years of schooling, are treated as exogenous. However, when we use instruments, the finite sample and asymptotic intervals differ significantly. In each of these examples, we also consider specifications that include only a constant and the covariate of interest. In these two dimensional situations, computation is relatively simple, so we consider estimating the finite sample intervals using a simple grid search, MCMC, and the marginal inference approach suggested in the previous section. We find that all methods result in similar confidence bounds for the parameter of interest in the demand example, but there are some discrepancies in the schooling example.

3.1. **Simulation Examples.** To illustrate the use of the asymptotic and finite sample approaches to inference, we conducted a series of simulation studies, the results of which are summarized in Table 1. In each panel of Table 1, the first row corresponds to testing the marginal hypothesis that $\theta(\tau)_{[1]} = \theta_0(\tau)_{[1]}$ where $\theta(\tau)_{[1]}$ is the first element of vector $\theta(\tau)$, and the second row corresponds to testing the joint hypothesis that $\theta(\tau) = \theta_0(\tau)$. For each model, we report inference results for the median, 75th percentile, and 90th percentile, i.e. $\tau \in \{.5, .75, .9\}$. The first three columns correspond to results obtained using the usual asymptotic approximation[6] and the last three columns correspond to results obtained via the finite sample approach. All results are for 5% level tests.

Panel A of Table 1 corresponds to a linear location-scale model with no endogeneity. The simulation model is given by $Y = D + (1 + D)\epsilon$ where $D \sim \text{BETA}(1, 1)$ and $\epsilon \sim N(0, 1)$. The sample size is 100. The conditional quantiles in this model are given by $q(D, \theta_0, \tau) =$

---

[5] In all examples, we set use the identity function for $g(\cdot)$.

[6] In the exogenous model, we base the asymptotic approximation on the conventional quantile regression of Koenker and Bassett (1978), using the Hall-Sheather bandwidth choice suggested by Koenker (2005), and we use the inverse quantile regression of Chernozhukov and Hansen (2001) in the endogenous settings, using the bandwidth choice suggested by Koenker (2005).

$\theta_0(\tau)_{[2]} + \theta_0(\tau)_{[1]}D$ where $\theta_0(\tau)_{[2]} = \Phi^{-1}(\tau)$ and $\theta_0(\tau)_{[1]} = 1 + \Phi^{-1}(\tau)$ for $\Phi^{-1}(\tau)$ the inverse of the normal CDF evaluated at τ.

Looking first at results for joint inference, we see that the finite sample procedure produces tests with the correct size at each of the three quantiles considered. On the other hand, the asymptotic approximation results in tests which overreject in each of the three cases, with the size distortion increasing as one moves toward the tail quantiles of the distribution. This behavior is unsurprising given that the usual asymptotically normal inference is inappropriate for inference about the tail quantiles of the distribution (see e.g. Chernozhukov (2000)) while the finite sample approach remains valid.

When we look at marginal inference about $\theta(\tau)_{[1]}$, we again see that tests based on the asymptotic distribution are size distorted with the distortion increasing as one moves toward the tail, though the distortions are smaller than for joint inference. Here, the finite sample inference continues to provide valid inference in that the size of the test is smaller than the nominal level. However, the finite sample approach appears to be quite conservative, rejecting far less frequently than the 5% level would suggest.

To further explore the conservativity of the finite sample approach, we plot power curves for tests based on the finite sample and usual asymptotic approximation in Figure 2. In this figure, values for $\theta_A(\tau)_{[1]} - \theta_0(\tau)_{[1]}$ where $\theta_A(\tau)_{[1]}$ is the hypothesized value for $\theta(\tau)_{[1]}$ are on the horizontal axis, and the vertical axis measures rejection frequencies of the hypothesis that $\theta(\tau)_{[1]} = \theta_A(\tau)_{[1]}$. Thus, size is given where the horizontal axis equals zero, and remaining points give power against various alternatives. The solid line in the figure gives the power curve for the test using the finite sample approach, and the dashed line gives the power curve for the test using the asymptotic approximation. From this figure, we can see that while conservative, tests based on the finite sample procedure do have some power against alternatives. The finite sample power curve always lies below the corresponding power curve generated from the asymptotic approximation, though this must be interpreted with caution due to the distortion in the asymptotic tests.

In Panels B-D of Table 1, we consider the performance of the asymptotic approximation and the finite sample inference approach in a model with endogeneity. The data for this simulation are generated from a location model with one endogenous regressor and three instrumental variables. In particular, we have $Y = -1 + D + \epsilon$ and $D = \Pi Z_1 + \Pi Z_2 + \Pi Z_3 + V$ where $Z_j \sim N(0,1)$ for $j = 1,2,3$, $\epsilon \sim N(0,1)$, $V \sim N(0,1)$, and the correlation between ϵ and v is .8. As above, this leads to a linear structural quantile function of the form $q(D, \theta_0, \tau) = \theta_0(\tau)_{[2]} + \theta_0(\tau)_{[1]}D$ where $\theta_0(\tau)_{[2]} = -1 + \Phi^{-1}(\tau)$ and $\theta_0(\tau)_{[1]} = 1$ for $\Phi^{-1}(\tau)$ the inverse of the normal CDF evaluated at τ. As above, the sample size is 100.

We explore the behavior of the inference procedures for differing degrees of correlation between the instruments and endogenous variable by changing Π across Panels B-D. In Panel B, we set $\Pi = 0.05$ which produces a first stage F-statistic of 0.205. In this case, the relationship between the instruments and endogenous variable is very weak and we would expect the asymptotic approximation to perform poorly. In Panel C, $\Pi = 0.5$ and the first stage F-statistic is 25.353. In Panel D, $\Pi = 1$ and the first stage F-statistic is 96.735. Both of these specifications correspond to fairly strong relationships between the endogenous variable and the instruments, and we would expect the asymptotic approximation to perform reasonably well in both cases. The finite sample procedure, on the other hand, should provide accurate inference in all three cases.

As expected, the tests based on the asymptotic approximation perform quite poorly in the weakly identified case presented in Panel B. Rejection frequencies for the asymptotic tests range from a minimum of .235 to a maximum of .474 for a 5% level test. In terms of size, the finite sample procedure performs quite well. As indicated by the theory, the finite sample approach has approximately the correct size for performing tests about the entire parameter vector. When considering the slope coefficient only, the finite sample procedure is conservative with rejection rates of .024 for $\tau = .5$, .0212 for $\tau = .75$, and .02 for $\tau = .9$ quantile.

The results for the models where identification is stronger given in Panels C and D are similar though not nearly so dramatic. The hypothesis tests based on the asymptotic procedure overreject in almost every case, with the lone exception being testing the joint hypothesis at the median when $\Pi = 1$. The size distortions increase as one moves from $\tau = .5$ to $\tau = .9$, and they decrease when Π increases from .5 to 1. The distortions at the 90th percentile remain quite large with rejection frequencies ranging between .1168 and .1540. For $\tau = .5$ and $\tau = .75$, the distortions are more modest with rejection rates ranging between .068 and .086.

The finite sample results are much more stable than the asymptotic results. The joint tests, with rejection frequencies ranging between 4.6% and 5.5%, do not appear to be size distorted. Marginal inference remains quite conservative with sizes ranging from .0198 to .0300. The results clearly suggest that the finite sample inference procedure is preferable for testing joint hypotheses, and given the size distortions found in the asymptotic approach the results also seem to favor the finite sample procedure for marginal inference.

As above, we also plot power curves for the asymptotic and finite sample testing procedures in Figures 3-5. Figure 3 contains power curves for $\Pi = .05$. In this case, the finite sample procedure appears to have essentially no power against any alternative. The lack of power is unsurprising given that identification in this case is extremely weak. In Figures 4 and 5, where the correlation between the instruments and endogenous regressors is stronger, the finite sample procedure seems to have quite reasonable power. The power curves for the finite sample

procedure are similar to the power curves of the asymptotic tests across a large portion of the parameter space. The finite sample procedure does have lower power against some alternatives that are near to the true parameter value than the asymptotic tests, though again this must be interpreted with some caution due to the distortions in the asymptotic tests.

Overall, the simulation results are quite favorable for the finite sample procedure. The results for joint inference confirm the theoretical properties of the procedure and suggest that numeric approximation error is not a large problem as the tests all have approximately correct size. For tests of joint hypotheses, the finite sample procedure clearly dominates the asymptotic procedure which may be substantially size distorted. For marginal inference, the results are somewhat less clear cut though still favorable for the finite sample procedure. In this case, the finite sample procedure may result in tests that are quite conservative, though the tests do appear to have nontrivial power against many hypotheses. On the other hand, tests based on the asymptotic approximation have size greater than the nominal level in the simulation models considered.

3.2. **Case Studies.** *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: $\ln(Q_p) = \alpha_0(U) + \alpha_1(U) \ln(p) + X'\beta(U)$, 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 a $U(0,1)$ distribution, $\alpha_1(U)$ is the random demand elasticity when the level of demand is U, and X is a vector of indicator variables for day of the week that enter the model with random coefficient $\beta(U)$. We consider two different specifications. In the first, we set $\beta(U) = 0$, and in the second, we estimate $\beta(U)$. 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.

As instruments, 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.[7] 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.

Asymptotic intervals are based on the inverse quantile regression estimator of Chernozhukov and Hansen (2001) when we treat price as endogenous. For models in which we set $D = Z$, i.e. in which we treat the covariates as exogenous, we base the asymptotic intervals on the conventional quantile regression estimator of Koenker and Bassett (1978).[8]

Estimation results are presented in Table 2. Panel A of Table 2 gives estimation results treating price as exogenous, and Panel B contains confidence intervals for the random elasticities when we instrument for price using both of the weather condition instruments described above. Panels C and D include a set of dummy variables for day of the week as additional covariates and are otherwise identical to Panels A and B respectively. In every case, we provide estimates of the 95%-level confidence interval obtained from the usual asymptotic approximation and the finite sample procedure. For the finite sample procedure, we report intervals obtained via MCMC, a grid search,[9] and the marginal procedure[10] in Panels A and B. In Panels C and D, we report only intervals constructed using the asymptotic approximation and the marginal procedure. For each model, we report estimates for $\tau = .25$, $\tau = .50$, and $\tau = .75$.

Looking first at Panels A and C which report results for models that treat price as exogenous, we see modest differences between the asymptotic and finite sample intervals. At the median when no covariates (other than price and intercept) are included, the asymptotic 95% level interval is (-0.785,-0.037), and the widest of the finite sample intervals is (-1.040,0.040). The differences become more pronounced at the 25th and 75th percentiles where we would expect the asymptotic approximation to be less accurate than at the center of the distribution. When day of the week effects are included, the asymptotic intervals tend to become narrower while the finite sample intervals widen slightly leading to larger differences in this case. However, the

---

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

[8]We use the Hall-Sheather bandwidth choice suggested by Koenker (2005) to implement the asymptotic standard errors.

[9]When price is treated as exogenous, we use an equally spaced grid over [5,10] with spacing .02 for α_0 and an equally spaced grid over [-4,2] with spacing .015 for α_1 for all quantiles. When price is treated as endogenous, we use different grid search regions for each quantile. For $\tau = .25$, we used an equally spaced grid over [0,10] with spacing .025 for α_0 and an equally spaced grid over [-40,40] with spacing .25 for α_1. For $\tau = .50$, we used an equally spaced grid over [6,12] with spacing .0125 for α_0 and an equally spaced grid over [-5,5] with spacing .025 for α_1. For $\tau = .75$, we used an equally spaced grid over [0,30] with spacing .05 for α_0 and an equally spaced grid over [-10,30] with spacing .05 for α_1.

[10]For the marginal procedure, we considered an equally spaced grid over [-5,1] at .01 unit intervals for all models.

basic results remain unchanged. Also, it is worth noting that all three computational methods for obtaining the finite sample confidence bounds give similar answers in the model with only an intercept and price with the marginal approach performing slightly better than the other two procedures. This finding provides some evidence that MCMC and the marginal approach may do as well computationally as a grid search which may not be feasible in high dimensional problems.

Turning now to results for estimation of the demand model using instrumental variables in Panels B and D, we see quite large differences between the asymptotic intervals and the intervals constructed using the finite sample approach. As above the differences are particularly pronounced at the 25th and 75th percentiles where the finite sample intervals are extremely wide. Even at the median in the model with only price and an intercept, the finite sample intervals are approximately twice as wide as the corresponding asymptotic intervals. When additional controls are included, the finite sample bounds for all three quantiles include the entire grid search region. The large differences between the finite sample and asymptotic intervals definitely call into question the validity of the asymptotic approximation in this case, which is not surprising given the relatively small sample size and the fact that we are estimating a nonlinear instrumental variables model.

Finally, it is worth noting again the three approaches to constructing the finite sample interval in general give similar results in this case. The differences between the grid search and marginal approaches could easily be resolved by increasing the search region for the marginal approach which was restricted to values we felt were a priori plausible. The difference between the grid search and MCMC intervals at the 25th percentile is more troubling, though it could likely be resolved through additional simulations or starting points.

As a final illustration, plots of 95% confidence regions in the model that includes only price and an intercept are provided in Figures 6 and 7. Figure 6 contains confidence regions for the coefficients treating price as exogenous, and Figure 7 contains confidence regions in the model where price is instrumented for using weather conditions. In the exogenous case, all of the regions are more or less elliptical and seem to be well-behaved. In this case, it is not surprising that all of the procedures for generating finite sample intervals produce similar results. The regions in Figure 7, on the other hand, are not nearly so well-behaved. In general, they are irregular and in many cases appear to be disconnected. The apparent failure of MCMC at the .25 quantile in the results in Table 2 is almost certainly due to the fact that the confidence region appears to be disconnected. The MCMC algorithm explores one of the regions but fails to jump to the other region. In cases like this, it is unlikely that a simple random walk Metropolis-Hastings algorithm will be sufficient to explore the space. While more complicated

MCMC or alternative stochastic search schemes could be explored, it seems that the marginal procedure is a convenient method to pursue if one is interested solely in marginal inference.

2. Returns to Schooling. As our final example, we consider estimation of a simple returns to schooling model that allows for heterogeneity in the effect of schooling on wages. We use data and the basic identification strategy employed in the schooling study of Angrist and Krueger (1991). The data are drawn from the 1980 U.S. Census and include observations on men born between 1930 and 1939. The data contain information on wages, years of completed schooling, state and year of birth, and quarter of birth. The total sample consists of 329,509 observations.

As in the previous section, we focus on a simple linear quantile model of the form $Y = \alpha_0(U) + \alpha_1(U)S + X'\beta(U)$ where Y is the log of the weekly wage, S is years of completed schooling, X is a vector 51 state of birth and 9 year of birth dummies that enter with random coefficients $\beta(U)$, and U is an unobservable normalized to follow a uniform distribution over $(0,1)$. We might think of U as indexing unobserved ability, in which case $\alpha_1(\tau)$ may be thought of as the return to schooling for an individual with unobserved ability τ. Since we believe that years of schooling may be jointly determined with unobserved ability, we use quarter of birth as an instrument for schooling, following Angrist and Krueger (1991). We consider two different specifications. In the first, we set $\beta(U) = 0$, and in the second, we estimate $\beta(U)$.

As in the previous example, we construct asymptotic intervals using the inverse quantile regression estimator when we treat schooling as endogenous. For models in which we treat schooling as exogenous, we construct the asymptotic intervals using the conventional quantile regression estimator.

We present estimation results in Table 3. Panel A of Table 3 gives estimation results treating schooling as exogenous, and Panel B contains confidence intervals for the schooling effect when we instrument for schooling using quarter of birth. Panels C and D include a set of 51 state of birth and 9 year of birth dummy variables but are otherwise identical to Panels A and B respectively. In every case, we provide estimates of the 95% confidence interval obtained from the usual asymptotic approximation and the finite sample procedure. For the finite sample procedure, we report intervals obtained via MCMC and a modified MCMC procedure (MCMC-2) that better accounts for the specifics of the problem, a grid search,[11] and the marginal procedure[12] in Panels A and B. The modified MCMC procedure we employ is a

---

[11]When price is treated as exogenous, we use unequally spaced grids over [4.4,5.6] for α_0 and unequally spaced grids over [0.062,0.077] for α_1 where the spacing depends on the quantile under consideration. When price is treated as endogenous, we use an equally spaced grid over [3,6] with spacing .01 for α_0 and an equally spaced grid over [0,0.25] with spacing .001 for α_1 for all quantiles.

[12]For the marginal procedure, we considered an equally spaced grid over [0.055,0.077] at .0004 unit intervals in the exogenous case for all models, and we used an equally spaced grid over [-1,1] at .01 unit intervals in the endogenous case.

simple stochastic search algorithm that simultaneously runs five MCMC chains each started at a local mode of the objective function. The idea behind the procedure is that the simple MCMC tends to get "stuck" because of the sharpness of the contours in this problem. By using multiple chains started at different values, we may potentially explore more of the function even if the chains get stuck near a local mode. If the starting points sufficiently cover the function, the approach should accurately recover the confidence region more quickly than the unadjusted MCMC procedure. In Panels C and D, we report only intervals constructed using the asymptotic approximation and the marginal procedure. For each model, we report estimates for $\tau = .25$, $\tau = .50$, and $\tau = .75$.

Looking first at estimates of the conditional quantiles of log wages given schooling presented in Panels A and C, we see that there is very little difference between the finite sample and asymptotic inference results. In Panel A where the model includes only a constant and the schooling variable, the finite sample and asymptotic intervals are almost identical. There are larger differences between the finite sample and asymptotic intervals in Panel C which includes 51 state of birth effects and 9 year of birth effects in addition to the schooling variable; though even in this case the differences are quite small. The close correspondence between the results in not surprising since in the exogenous case the parameters are well-identified and the sample is large enough that one would expect the asymptotic approximation to perform quite well for all but the most extreme quantiles.

While there is close agreement between the finite sample and asymptotic results in the model which treats schooling as exogenous, there are still substantial differences between the asymptotic and finite sample results in the case where we instrument for schooling using quarter of birth. The finite sample intervals, with the exception of the interval at the median, are substantially wider than the asymptotic intervals in the model with only schooling and an intercept, though in all cases they exclude zero. When we consider the finite sample intervals in the model that includes the state of birth and year of birth covariates, the differences are huge. For all three quantiles, the finite sample interval includes at least one endpoint of the search region, and in no case are the bounds informative. While the finite sample bounds may be quite conservative in models with covariates, the differences in this case are extreme. Also, we have evidence from the model which treats education as exogenous that in a well-identified setting the inflation of the bounds need not be large. Taken together, this suggests that identification in this model is quite weak.

While the finite sample intervals constructed through the different methods are similar at the median in the instrumented model, there are large differences between the finite sample intervals for the .25 and .75 quantiles with the simple MCMC performing the worst and the marginal approach performing the best. The difficulty in this case is that the objective function

has extremely sharp contours. This sharpness of contours is illustrated in Figure 8 which plots the 95% level confidence region obtained from the MCMC-2 procedure for the .75 quantile in the schooling example without covariates.

The shape of the confidence region poses difficulties for both the traditional grid search and the basic MCMC procedure. The problem with the grid search is that the interval is so narrow that even with a very fine grid one is unlikely to find more than a few points in the region unless the grid is chosen carefully to include many points along the "line" describing the confidence region, and with a course grid, one may miss the confidence region entirely. The narrowness of the confidence set also causes problems with MCMC by making transitions quite difficult. Essentially with a default random walk Metropolis-Hastings procedure one must specify either a very small variance for the transition density or must specify the correlation exactly so that the proposals lie along the "line" describing the contours. Designing a transition density with the appropriate covariance is complicated as even slight perturbations may result in proposals that lie off of the line making transitions unlikely unless the variance is small. Taken together this suggests that MCMC is likely to travel very slowly through the parameter space resulting in poor convergence properties and difficulty in generating the finite sample confidence regions.

The MCMC-2 procedure alleviates the problems with the random walk MCMC somewhat by running multiple chains with different starting values. Using multiple chains provides local exploration of the objective function around the starting values. In cases where the objective function is largely concentrated around a few local modes, this provides improvement in generating the finite sample confidence regions. This approach is still insufficient in this example at the .25 quantile where we see that the MCMC-2 interval is still significantly shorter than the interval generated through the marginal approach suggesting that the MCMC-2 procedure did not travel sufficiently through the parameter space.

In this example, the marginal approach seems to clearly dominate the other approaches to computing the finite sample confidence regions that we have considered. It finds more points that lie within the confidence bound for the parameter of interest than any of the other approaches. It is also simple to implement, and the search region can be chosen to produce a desired level of accuracy.

4. Conclusion

In this paper, we have presented an approach to inference in models defined by quantile restrictions that is valid under minimal assumptions. The approach does not rely on any asymptotic arguments, does not require the imposition of distributional assumptions, and will be valid for both linear and nonlinear conditional quantile models and for models which include endogenous as well as exogenous variables. The approach relies on the fact that objective

functions that quantile regression aims to solve are conditionally pivotal in finite samples. This conditional pivotal property allows the construction of exact finite sample joint confidence regions and on finite sample confidence bounds for quantile regression coefficients.

The chief drawbacks of the approach are that it may be computationally difficult and that it may be quite conservative for performing inference about subsets of regression parameters. We suggest that MCMC or other stochastic search algorithms may be used to construct joint confidence regions. In addition, we suggest a simple algorithm that combines optimization with a one-dimensional search that can be used to construct confidence bounds for individual regression parameters. In simulations, we find that the finite sample inference procedure is not conservative for testing hypotheses about the entire vector of regression parameters but that it is conservative for tests about individual regression parameters. However, the finite sample tests do have moderate power in many situations, and tests based on the asymptotic approximation tend to overreject. Overall, the findings of the simulation study are quite favorable to the finite sample approach.

We also consider the use of the finite sample inference in two simple empirical examples: estimation of a demand curve in a small sample and estimation of the returns to schooling in a large sample. In the demand example, we find modest differences between the finite sample and asymptotic intervals when we estimate conditional quantiles not instrumenting for price and large differences when we instrument for price. In the schooling example, the finite sample and asymptotic intervals are almost identical in models in which we treat schooling as exogenous, and again there are large differences in the approaches when we instrument for schooling. These results suggest that in both cases, the identification of the structural parameters in the instrumental variables models is weak.

Appendix A. Appendix: Optimality Arguments for L_n.

In the preceding sections, we introduced a finite sample inference procedure for quantile regression models and demonstrated that this procedure provides valid inference statements in finite samples. In this section, we show that the approach also has desirable large sample properties:

(1) Under strong identification, the class of statistics of the form (2.3) contains a (locally) asymptotically uniformly most powerful (UMP) invariant test. Inversion of this test therefore gives (locally) uniformly most accurate invariant regions. (The definitions of power and invariance follow those in Choi, Hall, and Schick (1996)).

(2) Under weak identification, the class of statistics of the form (2.3) maximizes an average power function within a broad class of normal weight functions.

Here, we suppose $(Y_i, D_i, Z_i, i = 1, \ldots, n)$ is an i.i.d. sample from the model defined by A1-6 and assume that the dimension K of θ_0 is fixed. Although this assumption can be relaxed, the primary

purpose of this section is to motivate the statistics used for finite-sample inference from an optimality point of view.

Recall that under A1-6, $P\left[Y - q\left(D, \theta_0, \tau\right) \leq 0 | Z\right] = \tau$. Consider the problem of testing

$$H_0 : \theta_0 = \theta_* \qquad \text{vs.} \qquad H_a : \theta_0 \neq \theta_*,$$

where $\theta_* \in \mathbb{R}^K$ is some constant.

Let $e_i = 1\left[Y_i \leq q\left(D_i, \theta_*, \tau\right)\right]$. As defined, $e|Z \sim \text{Bernoulli}[\tau\left(Z, \theta_0\right)]$, where $\tau\left(Z, \theta_0\right) = P\left[Y \leq q\left(D, \theta_0, \tau\right) | Z\right]$. Suppose testing is to be based on $\left(e_i, Z_i, i = 1, \ldots, n\right)$. Because $e_i | Z_1, \ldots, Z_n \sim i.i.d. \text{Bernoulli}(\tau)$ under the null, any statistic based on $\left(e_i, Z_i, i = 1, \ldots, n\right)$ is conditionally pivotal under H_0.

Let $\mathcal{G}$ be the class of functions g for which $E\left[g\left(Z\right) g\left(Z\right)'\right]$ exists and is positive definite; that is, let $\mathcal{G} = \cup_{j=1}^{\infty} \mathcal{G}_j$, where $\mathcal{G}_j$ is the class of $\mathbb{R}^j$-valued functions g for which $E\left[g\left(Z\right) g\left(Z\right)'\right]$ exists and is positive definite. As mentioned in the text, a "natural" class of test statistics is given by $\left\{L_n\left(\theta_*, g\right) : g \in \mathcal{G}\right\}$, where

$$L_n\left(\theta_*, g\right) = \left[\sum_{i=1}^{n} g\left(Z_i\right)\left(e_i - \tau\right)\right]' \left[\tau\left(1 - \tau\right) \sum_{i=1}^{n} g\left(Z_i\right) g\left(Z_i\right)'\right]^{-1} \left[\sum_{i=1}^{n} g\left(Z_i\right)\left(e_i - \tau\right)\right]. \qquad \text{(A.1)}$$

Being based on $\left(e_i, Z_i, i = 1, \ldots, n\right)$, any such $L_n\left(\theta_*, g\right)$ is conditionally pivotal under the null. In addition, under the null,

$$L_n\left(\theta_*, g\right) \rightarrow_d \frac{1}{2} \chi^2_{\dim(g)}$$

for any $g \in \mathcal{G}$. Moreover, the class $\left\{L_n\left(\theta_*, g\right) : g \in \mathcal{G}\right\}$ enjoys desirable large sample power properties under the following strong identification assumption in which Θ_* denotes some open neighborhood of θ_*.

Assumption 3. (a) The distribution of Z does not depend on θ_0. (b) For every $\theta \in \Theta_*$ (and for almost every Z),

$$\dot{\tau}\left(Z, \theta\right) = \frac{\partial}{\partial \theta} \tau\left(Z, \theta\right) \qquad \text{(A.2)}$$

exists and is continuous (in θ). (c) $\dot{\tau}_*\left(Z\right) = \dot{\tau}\left(Z, \theta_*\right) \in \mathcal{G}$.(d) $E\left[\sup_{\theta \in \Theta_*} \|\dot{\tau}\left(Z, \theta\right)\|^2\right] < \infty$.

If Assumption 3 holds and $g \in \mathcal{G}$, then under contiguous alternatives induced by the sequence $\theta_{0,n} = \theta_* + b/\sqrt{n}$,

$$L_n\left(\theta_*, g\right) \rightarrow_d \frac{1}{2} \chi^2_{\dim(g)}\left[\frac{1}{\tau\left(1 - \tau\right)} \delta_S\left(b, g\right)\right], \qquad \text{(A.3)}$$

where

$$\delta_S\left(b, g\right) = b' E\left[\dot{\tau}_*\left(Z\right) g\left(Z\right)'\right] E\left[g\left(Z\right) g\left(Z\right)'\right]^{-1} E\left[g\left(Z\right) \dot{\tau}_*\left(Z\right)'\right] b.$$

By a standard argument, $\delta_S\left(b, g\right) \leq \delta_S\left(b, \dot{\tau}_*\right)$ for any $g \in \mathcal{G}$. As a consequence, $L_n\left(\theta_*, \dot{\tau}_*\right)$ maximizes local asymptotic asymptotic power within the class $\left\{L_n\left(\theta_*, g\right) : g \in \mathcal{G}\right\}$. An even stronger optimality result is the following.

Proposition 4. *Among tests based on* $(e_i, Z_i, i = 1, \ldots, n)$, *the test which rejects for large values of* $L_n(\theta_*, \dot{\tau}_*)$ *is a locally asymptotically UMP (rotation) invariant test of* H_0. *Therefore,* $\{L_n(\theta_*, g) : g \in \mathcal{G}\}$ *is an (asymptotically) essentially complete class of tests of* H_0 *under Assumption 3.*

PROOF: The conditional (on $\underline{Z} = (Z_1, \ldots, Z_n)$) log likelihood function is given by

$$\ell_n(\theta|\underline{Z}) = \sum_{i=1}^{n} \{\log[\tau(Z_i, \theta)] e_i + \log[1 - \tau(Z_i, \theta)](1 - e_i)\}.$$

Assumption 3 implies that the following LAN expansion is valid under the null: For any $b \in \mathbb{R}^K$,

$$\ell_n\left(\theta_* + \frac{b}{\sqrt{n}}\right) - \ell_n(\theta_*) = b'\mathcal{S}_n^* - \frac{1}{2}b'\mathcal{I}_n^* b + o_p(1),$$

where ℓ_n is the (unconditional) log likelihood function,

$$\mathcal{S}_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\tau(1-\tau)} \dot{\tau}_*(Z_i)(e_i - \tau) \to_d \mathcal{N}(0, \mathcal{I}^*),$$

and

$$\mathcal{I}_n^* = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\tau(1-\tau)} \dot{\tau}_*(Z_i) \dot{\tau}_*(Z_i)' \to_p \mathcal{I}^* = \frac{1}{\tau(1-\tau)} E\left[\dot{\tau}_*(Z) \dot{\tau}_*(Z)'\right].$$

Theorem 3 of Choi, Hall, and Schick (1996) now shows that $L_n(\theta_*, \dot{\tau}_*) = \frac{1}{2}\mathcal{S}_n^{*'}\mathcal{I}_n^{*-1}\mathcal{S}_n^*$ is the asymptotically UMP invariant test of H_0. $\qquad\square$

In view of Proposition 4, a key role is played by $\dot{\tau}_*$. This gradient will typically be unknown but will be estimable under various assumptions ranging from parametric assumptions to nonparametric ones. As an illustration, consider the linear quantile model

$$Y = D'\theta_0 + \varepsilon, \tag{A.4}$$

where $P[\varepsilon \leq 0|Z] = \tau$. If the conditional distribution of ε given (X, Z) admits a density (with respect to Lebesgue measure) $f_{\varepsilon|X,Z}(\cdot|X, Z)$ and certain additional mild conditions hold, then Assumption 3 is satisfied with $\dot{\tau}_*(Z) = -E\left[Df_{\varepsilon|X,Z}(0|X, Z)|Z\right]$, an object which can be estimated nonparametrically. If, moreover, it is assumed that

$$D = \Pi'Z + v, \tag{A.5}$$

where $(\varepsilon, v')|Z \sim \mathcal{N}(0, \Sigma)$ for some positive definite matrix Σ, then $\dot{\tau}_*(Z)$ is proportional to $\Pi'Z$ and parametric estimation of $\dot{\tau}_*$ becomes feasible. (Assuming that the gradient belongs to a particular subclass of $\mathcal{G}$ will not affect the optimality result, as Proposition 4 (tacitly) assume that $\dot{\tau}_*$ is known.) Estimation of the gradient will not affect the asymptotic validity of the test even if the full sample is used, nor will it affect the validity of finite-sample inference provided sample splitting is used (i.e., estimation of $\dot{\tau}_*$ and finite-sample inference are performed using different subsamples of the full sample).

Under weak identification, Proposition 4 will not hold as stated, but a closely related optimality result is available. The key difference between the strongly and weakly identified cases is that the defining property of a weakly identified model is that the counterpart of the gradient $\dot{\tau}_*$ is not consistently estimable. As such, asymptotic optimality results are too optimistic. Nevertheless, it is still possible to

24

show the statistic used in the main text has an attractive optimality property under the following weak identification assumption in which $\tau(Z, \theta_0)$ is modeled as a "locally linear" sequence of parameters.[13]

Assumption 4. (a) The distribution of Z does not depend on θ_0, (b) $\tau(Z, \theta_*) = \tau + n^{-1/2}[Z'C\Delta_\theta + R_n(Z, \theta_*, C)]$ for some $C \in \mathbb{R}^{\dim(Z) \times K}$ and some function R_n, where $\Delta_\theta = \theta_0 - \theta_*$. (c) $\Sigma_{ZZ} = E(ZZ')$ exists and is positive definite, (d) $\lim_{n \to \infty} E\left[R_n(Z, \theta, C)^2\right] = 0$ for every θ and every C.

If Assumption 4 holds and $g \in \mathcal{G}$, then

$$L_n(\theta_*, g) \to_d \frac{1}{2} \chi^2_{\dim(g)} \left[\frac{1}{\tau(1-\tau)} \delta_W(\Delta_\theta, C, g) \right], \tag{A.6}$$

where

$$\delta_W(\Delta_\theta, C, g) = \Delta_\theta' E\left[C'Zg(Z)'\right] E\left[g(Z)g(Z)'\right]^{-1} E\left[g(Z)Z'C\right] \Delta_\theta.$$

As in the strongly identified case, the limiting distribution of $L_n(\theta_*, g)$ is $\frac{1}{2}$ times a noncentral $\chi^2_{\dim(g)}$ in the weakly identified case. Within the class of tests based on a member of $\{L_n(\theta_*, g) : g \in \mathcal{G}\}$, the asymptotically most powerful test is the one based on $L_n(\theta_*, g_C)$, where $g_C(Z) = C'Z$. This test furthermore enjoys an optimality property analogous to the one established in Proposition 4. The proof of the result for $L_n(\theta_*, g_C)$ is identical to that of Proposition 4, with b, $\mathcal{S}_n^*$, and $\mathcal{I}_n^*$ of the latter proof replaced by Δ_θ,

$$\mathcal{S}_n(C) = C'\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{\tau(1-\tau)} Z_i(e_i - \tau) \right],$$

and

$$\mathcal{I}_n(C) = C'\left[\frac{1}{n} \sum_{i=1}^n \frac{1}{\tau(1-\tau)} Z_i Z_i' \right] C,$$

respectively. (In particular, the proof utilizes the fact that if C is known, then the statistic $\mathcal{S}_n(C)$ is asymptotically sufficient under Assumption 4.)

However, the consistent estimation of C is infeasible in the present (weakly identified) case. Indeed, because C cannot be treated "as if" it was known, it seems more reasonable to search for a test which is implementable without knowledge of C and enjoys an optimality property that does not rely on this knowledge. To that end, let

$$L_n^* = \left[\sum_{i=1}^n Z_i(e_i - \tau) \right]' \left[\tau(1-\tau) \sum_{i=1}^n Z_i Z_i' \right]^{-1} \left[\sum_{i=1}^n Z_i(e_i - \tau) \right]; \tag{A.7}$$

that is, let L_n^* be the particular member of $\{L_n(\theta_*, g) : g \in \mathcal{G}\}$ for which g is the identity mapping.

It follows from Muirhead (1982, Exercise 3.15 (d)) that for any $\kappa > 0$ and any $\dim(D) \times \dim(D)$ matrix Σ_{vv}, L_n^* is a strictly increasing transformation of

$$\int \exp\left(\frac{\kappa}{1+\kappa} L_n(\theta_*, g_C) \right) dJ(C; \Sigma_{vv}), \tag{A.8}$$

---

[13]Assumption 4 is motivated by the Gaussian model (A.4)-(A.5). In that model, parts (b) and (d) hold (with C proportional to $\sqrt{n}\Pi$) if part (c) does and Π varies with n in such a way that $\sqrt{n}\Pi$ is a constant $\dim(Z) \times K$ matrix (as in Staiger and Stock (1997)).

where $J(\cdot)$ is the cdf of the normal distribution with mean 0 and variance $\Sigma_{vv} \otimes \left(n^{-1}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1}$. In (A.8), the functional form of $J(\cdot)$ is "natural" insofar as it is corresponds to the weak instruments prior employed by Chamberlain and Imbens (2004). Moreover, following Andrews and Ploberger (1994), the integrand in (A.8) is obtained by averaging the LAN approximation to the likelihood ratio with respect to the weight/prior measure $K_C(\theta_0)$ associated with the distributional assumption $\Delta_\theta \sim \mathcal{N}\left[0, \kappa\mathcal{I}_n(C)^{-1}\right]$. In view of the foregoing, it follows that the statistic L_n^* enjoys weighted average power optimality properties of the Andrews and Ploberger (1994) variety.[14] This statement is formalized in the following result.

Proposition 5. *Among tests based on* $(e_i, Z_i, i = 1, \ldots, n)$*, under Assumption 4 the test based on* L_n^* *is asymptotically equivalent to the test that maximizes the asymptotic average power:*

$$\limsup_{n \to \infty} \int \int Pr(reject\ \theta^* | \theta_0, C) dK_C(\theta_0) dJ(C).$$

REFERENCES

ABADIE, A. (1995): "Changes in Spanish Labor Income Structure During the 1980s: A Quantile Regression Approach," CEMFI Working Paper No. 9521.

ANGRIST, J. D., AND A. B. KRUEGER (1991): "Does compulsory school attendance affect schooling and earnings," *Quarterly Journal of Economics*, 106, 979–1014.

BUCHINSKY, M. (1994): "Changes in U.S. wage structure 1963-87: An application of quantile regression," *Econometrica*, 62, 405–458.

CHAMBERLAIN, G. (1994): "Quantile regression censoring and the structure of wages," in *Advances in Econometrics*, ed. by C. Sims. Elsevier: New York.

CHERNOZHUKOV, V. (2000): "Extremal Quantile Regression (Linear Models and an Asymptotic Theory)," MIT, Department of Economics, Technical Report, published in Ann. Statist. 33, no. 2 (2005), pp. 806–839, download at www.mit.edu/~vchern.

CHERNOZHUKOV, V., AND C. HANSEN (2001): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics* (forthcoming).

———— (2005a): "Instrumental Variable Quantile Regression," mimeo.

———— (2005b): "An IV Model of Quantile Treatment Effects," *Econometrica*.

CHERNOZHUKOV, V., AND H. HONG (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.

CHERNOZHUKOV, V., G. IMBENS, AND W. NEWEY (2006): "Instrumental Variable Estimation of Nonseparable Models," forthcoming *Journal of Econometrics*.

CHESHER, A. (2003): "Identification in Nonseparable Models," forthcoming in *Econometrica*.

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., J. JUREČKOVÁ, R. KOENKER, AND S. PORTNOY (1993): "Tests of linear hypotheses based on regression rank scores," *J. Nonparametr. Statist.*, 2(4), 307–331.

---

[14]As discussed by Andrews and Ploberger (1994, p. 1384), L_n^* can therefore be interpreted as (being asymptotically equivalent to) a Bayesian posterior odds ratio.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14(3), 262–280.

HE, X., AND F. HU (2002): "Markov chain marginal bootstrap," *J. Amer. Statist. Assoc.*, 97(459), 783–795.

HU, L. (2002): "Estimation of a Censored Dynamic Panel Data Model," *Econometrica*, 70(6), 2499–2517.

HUBER, P. J. (1973): "Robust regression: asymptotics, conjectures and Monte Carlo," *Ann. Statist.*, 1, 799–821.

KOENKER, R. (1997): "Rank tests for linear models," in *Robust inference*, vol. 15 of *Handbook of Statist.*, pp. 175–199. North-Holland, Amsterdam.

———— (2005): *Quantile Regression.* Cambridge University Press.

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

KOENKER, R., AND Z. XIAO (2004a): "Quantile autoregression," *www.econ.uiuc.edu.*

———— (2004b): "Unit root quantile autoregression inference," *J. Amer. Statist. Assoc.*, 99(467), 775–787.

MACKINNON, W. J. (1964): "Table for both the sign test and distribution-free confidence intevals of the median for sample sizes to 1,000," *J. Amer. Statist. Assoc.*, 59, 935–956.

MATZKIN, R. L. (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71(5), 1339–1375.

NEWEY, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *J. Econometrics*, 79(1), 147–168.

PAKES, A., AND D. POLLARD (1989): "Simulation and Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

PARZEN, M. I., L. J. WEI, AND Z. YING (1994): "A resampling method based on pivotal estimating functions," *Biometrika*, 81(2), 341–350.

PORTNOY, S. (1985): "Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation," *Ann. Statist.*, 13(4), 1403–1417.

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

ROBERT, C. P., AND G. CASELLA (1998): *Monte Carlo Statistical Methods.* Springer-Verlag.

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with weak identification," *Econometrica*, 68(5), 1055–1096.

TAMER, E. (2003): "Incomplete simultaneous discrete response model with multiple equilibria," *Rev. Econom. Stud.*, 70(1), 147–165.

WALSH, J. E. (1960): "Nonparametric tests for median by interpolation from sign tests," *Ann. Inst. Statist. Math.*, 11, 183–188.
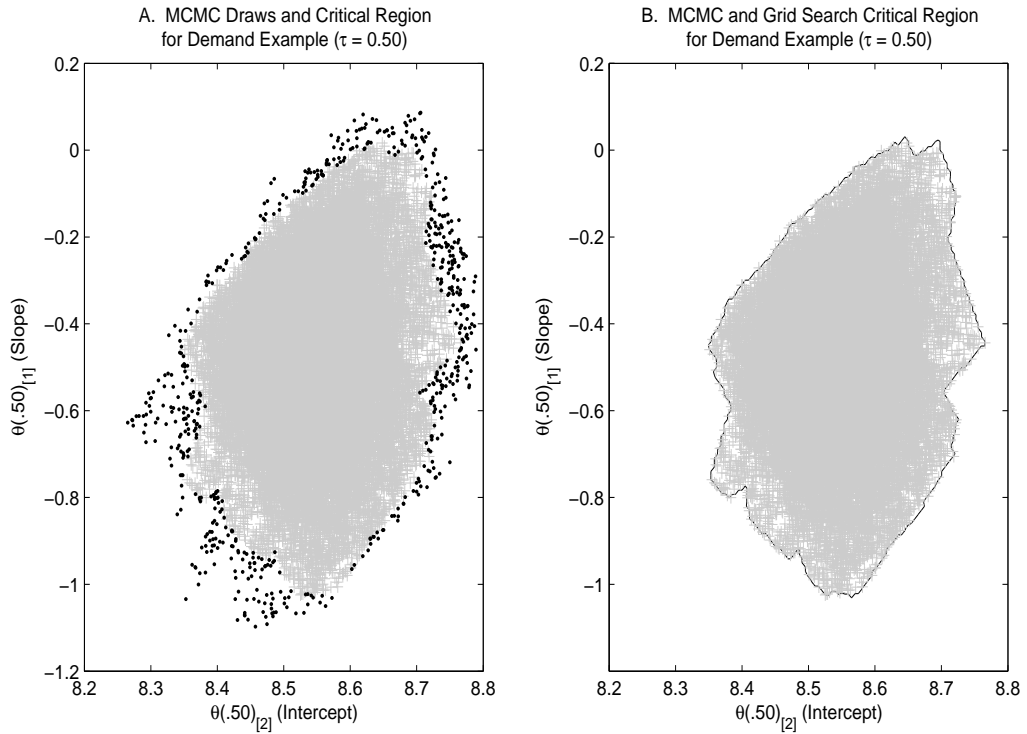
FIGURE 1. **MCMC and Grid Search Confidence Regions.** This figure illustrates the construction of a 95% level confidence regions by MCMC and a grid search in the demand example from Section 3. Panel A shows the MCMC draws. The gray +'s represent draws that fell within the confidence region, and the black •'s represent draws outside of the confidence region. Panel B plots the MCMC draws within the confidence region (gray +'s) and the grid search confidence region (black line).

<div align="center">Table 1. Monte Carlo Results</div>

Null Hypothesis	Asymptotic Inference			Finite Sample Inference		
	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
A. Exogenous Model						
$\theta(\tau)_{[1]} = \theta_0(\tau)_{[1]}$	0.0716	0.0676	0.0920	0.0080	0.0064	0.0044
$\theta(\tau) = \theta_0(\tau)$	0.0744	0.0820	0.1392	0.0516	0.0448	0.0448
B. Endogenous Model. $\Pi = .05$						
$\theta(\tau)_{[1]} = \theta_0(\tau)_{[1]}$	0.2380	0.2392	0.2720	0.0240	0.0212	0.0200
$\theta(\tau) = \theta_0(\tau)$	0.2352	0.3604	0.4740	0.0488	0.0460	0.0484
C. Endogenous Model. $\Pi = .5$						
$\theta(\tau)_{[1]} = \theta_0(\tau)_{[1]}$	0.0744	0.0808	0.1240	0.0300	0.0300	0.0204
$\theta(\tau) = \theta_0(\tau)$	0.0732	0.0860	0.1504	0.0552	0.0560	0.0464
D. Endogenous Model. $\Pi = 1$						
$\theta(\tau)_{[1]} = \theta_0(\tau)_{[1]}$	0.0632	0.0784	0.1168	0.0232	0.0216	0.0192
$\theta(\tau) = \theta_0(\tau)$	0.0508	0.0772	0.1440	0.0524	0.0476	0.0508

Note: Simulation results for asymptotic and finite sample inference for quantile regression. Each panel reports results for a different simulation model. Each simulation model has quantiles of the form $q(D, \theta_0, \tau) = \theta_0(\tau)_{[2]} + \theta_0(\tau)_{[1]}D$. The first row within each panel reports rejection frequencies for 5% level tests of the hypothesis that $\theta(\tau)_{[1]} = \theta_0(\tau)_{[1]}$, and the second row reports rejection frequencies for 5% level tests of the joint hypothesis $\theta = \theta_0$. The number of simulations is 2500.
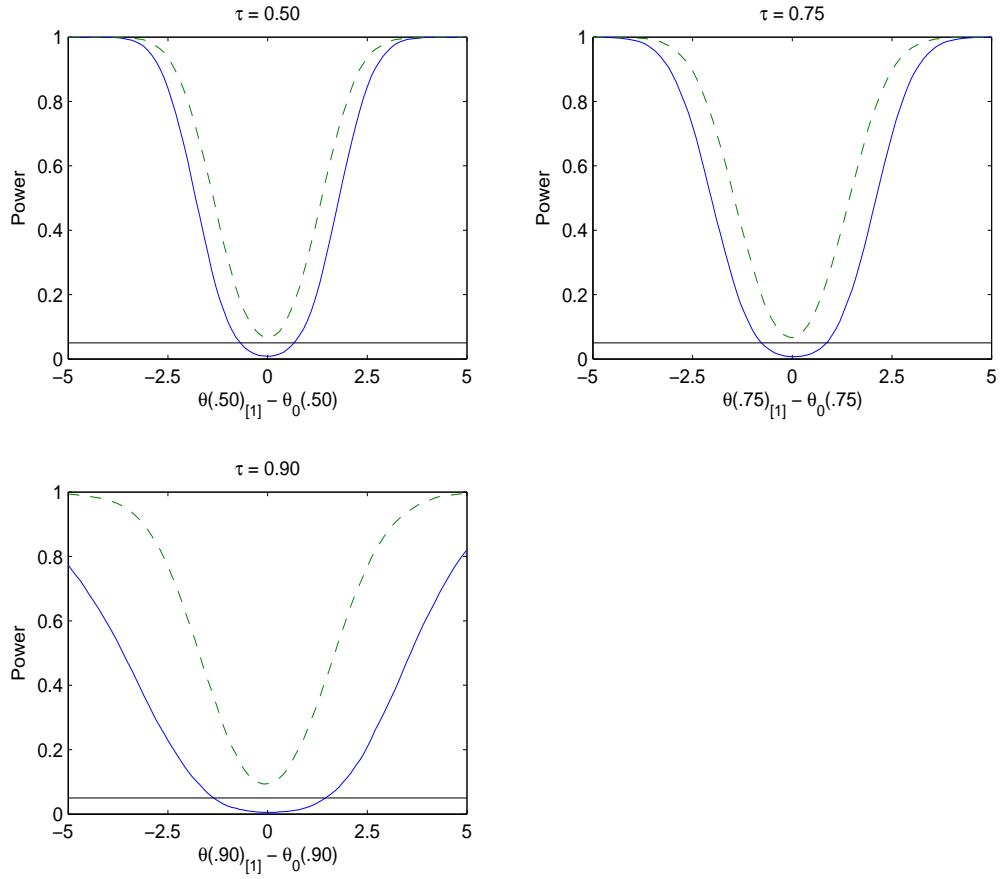
FIGURE 2. **Power Curves for Exogenous Simulation Model.** This figure plots power curves for the simulations contained in Panel A of Table 1. The solid line is the power curve for a test based on the finite sample inference procedure, and the dashed line is the power curve from a test based on the asymptotic approximation.
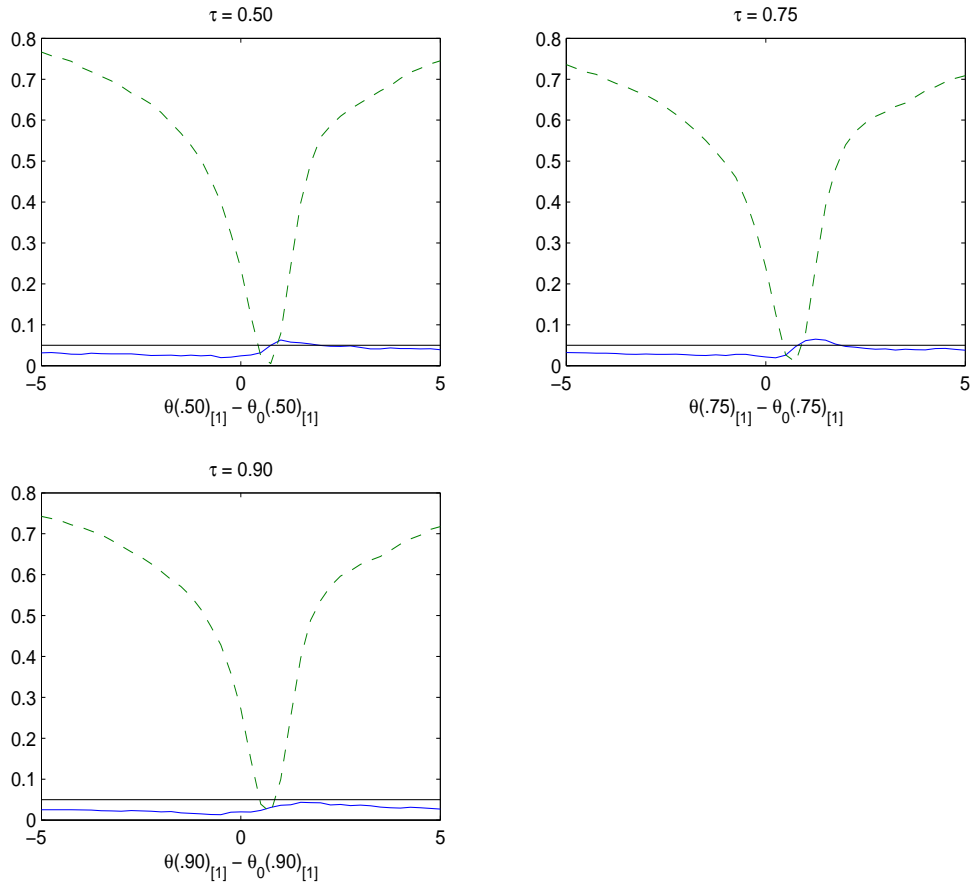
FIGURE 3. **Power Curves for Endogenous Simulation Model, $\Pi = 0.05$.**
This figure plots power curves for the simulations contained in Panel B of Table
1 which correspond to a nearly unidentified case. The solid line is the power
curve for a test based on the finite sample inference procedure, and the dashed
line is the power curve from a test based on the asymptotic approximation.
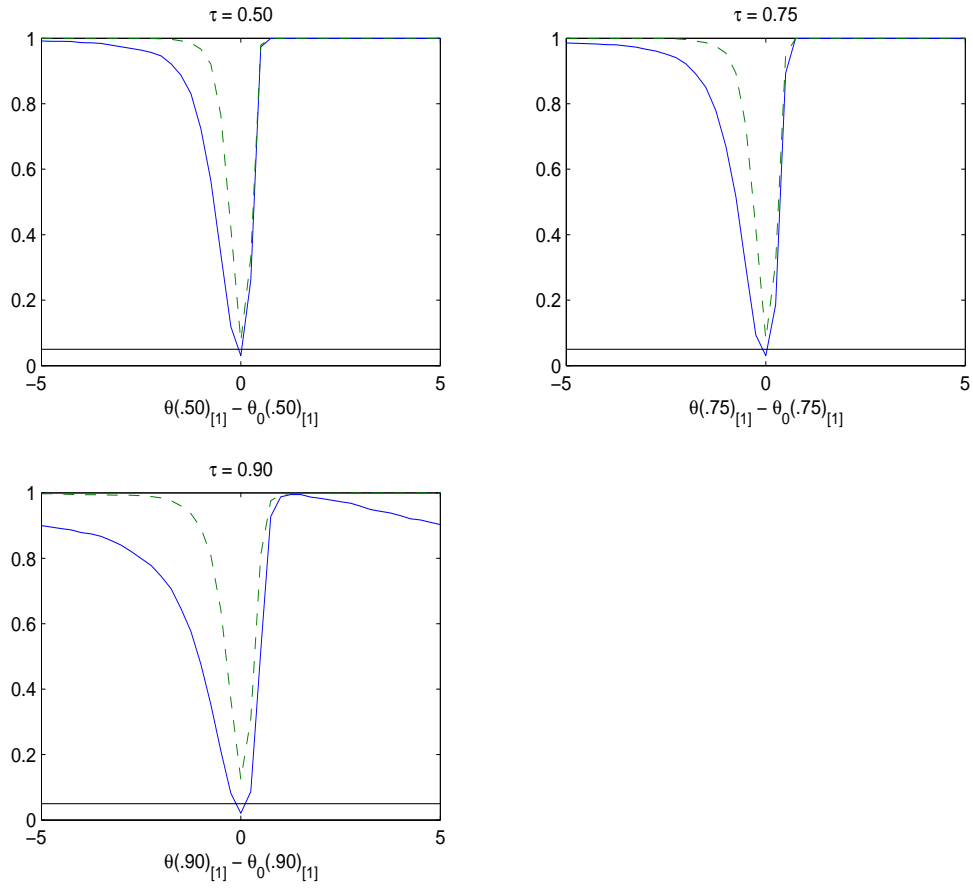
FIGURE 4. **Power Curves for Endogenous Simulation Model,** $\Pi = .5$**.**
This figure plots power curves for the simulations contained in Panel C of Table
1. The solid line is the power curve for a test based on the finite sample inference
procedure, and the dashed line is the power curve from a test based on the
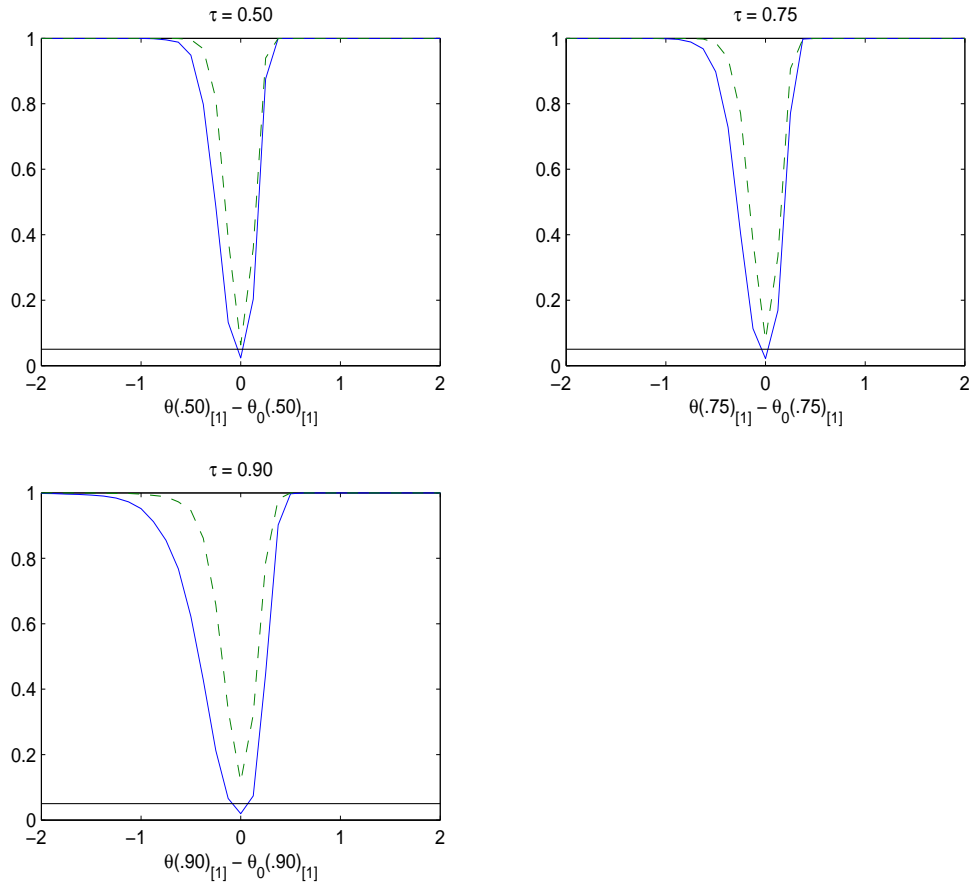asymptotic approximation.

FIGURE 5. **Power Curves for Endogenous Simulation Model, $\Pi = 1$.** This figure plots power curves for the simulations contained in Panel D of Table 1. The solid line is the power curve for a test based on the finite sample inference procedure, and the dashed line is the power curve from a test based on the asymptotic approximation.

TABLE 2. Demand for Fish

Estimation Method	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
A. Quantile Regression (No Instruments)			
Quantile Regression (Asymptotic)	(-0.874,0.073)	(-0.785,-0.037)	(-1.174,-0.242)
Finite Sample (MCMC)	(-1.348,0.338)	(-1.025,0.017)	(-1.198,0.085)
Finite Sample (Grid)	(-1.375,0.320)	(-1.015,0.020)	(-1.195,0.065)
Finite Sample (Marginal)	(-1.390,0.350)	(-1.040,0.040)	(-1.210,0.090)
B. IV Quantile Regression (Stormy, Mixed as Instruments)			
Inverse Quantile Regression (Asymptotic)	(-2.486,-0.250)	(-1.802,0.030)	(-2.035,-0.502)
Finite Sample (MCMC)	(-4.403,1.337)	(-3.566,0.166)	(-5.198,25.173)
Finite Sample (Grid)	(-4.250,40]	(-3.600,0.200)	(-5.150,24.850)
Finite Sample (Marginal)	(-4.430,1]	(-3.610,0.220)	[-5,1]
C. Quantile Regression - Day Effects (No Instruments)			
Quantile Regression (Asymptotic)	(-0.695,-0.016)	(-0.718,-0.058)	(-1.265,-0.329)
Finite Sample (Marginal)	(-1.610,0.580)	(-1.360,0.320)	(-1.350,0.400)
D. IV Quantile Regression - Day Effects (Stormy, Mixed as Instruments)			
Inverse Quantile Regression (Asymptotic)	(-2.403,-0.324)	(-1.457,0.267)	(-1.895,-0.463)
Finite Sample (Marginal)	[-5,1]	[-5,1]	[-5,1]

Note: 95% level confidence interval estimates for Demand for Fish example. Panel A reports results from model which treats price as exogenous, and Panel B reports results from model which treats price as endogenous and uses weather conditions as instruments for price. Panels C and D are as A and B but include a set of dummy variables for day of the week. The first row in each panel reports the interval estimated using the asymptotic approximation, and the remaining rows report estimates of the finite sample interval constructed through various methods.
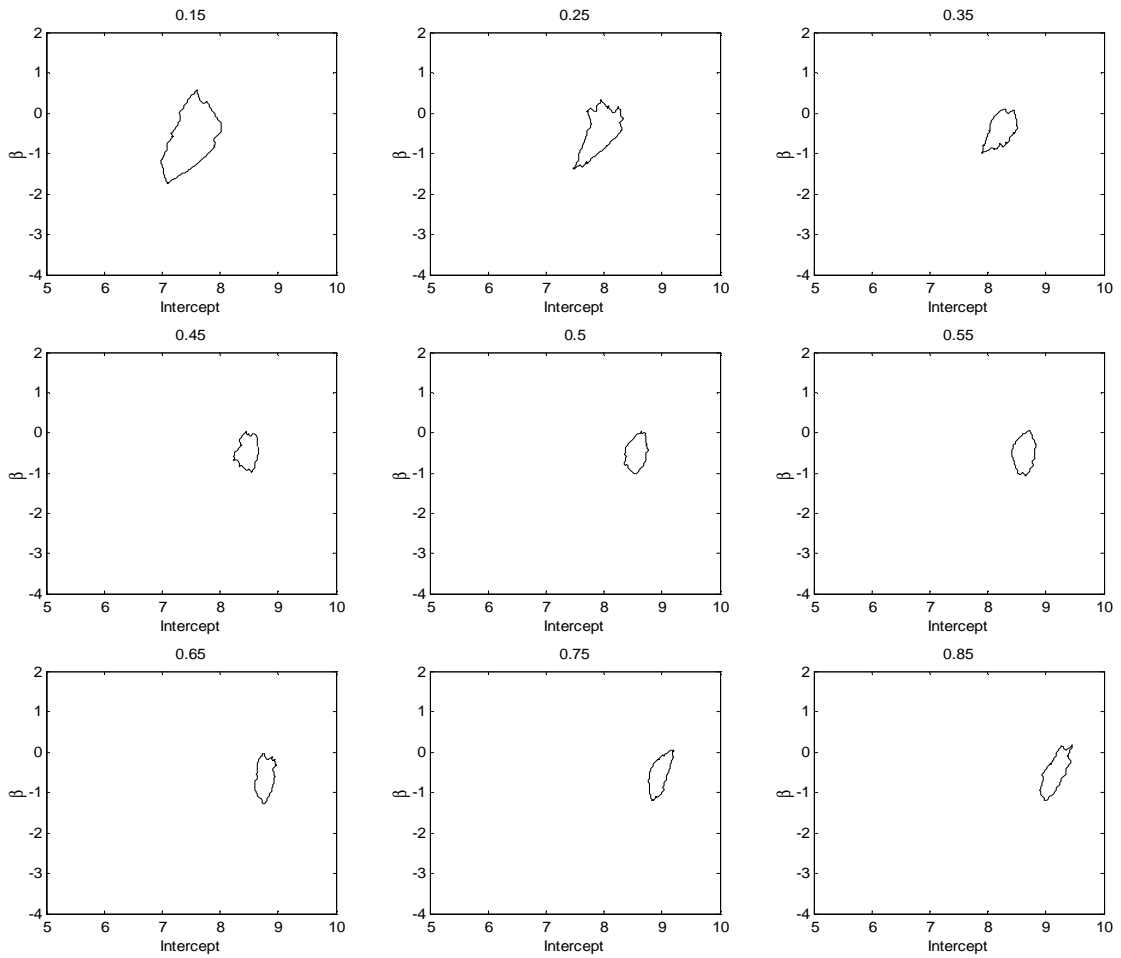
FIGURE 6. **Finite Sample Confidence Regions for Fish Example Treating Price as Exogenous.** This figure plots finite sample confidence regions from fish example without covariates treating price as exogenous. Values for the intercept, $\theta(\tau)_{[2]}$ are on the horizontal axis, and values for the slope parameter $\theta(\tau)_{[1]}$ are on the vertical axis.
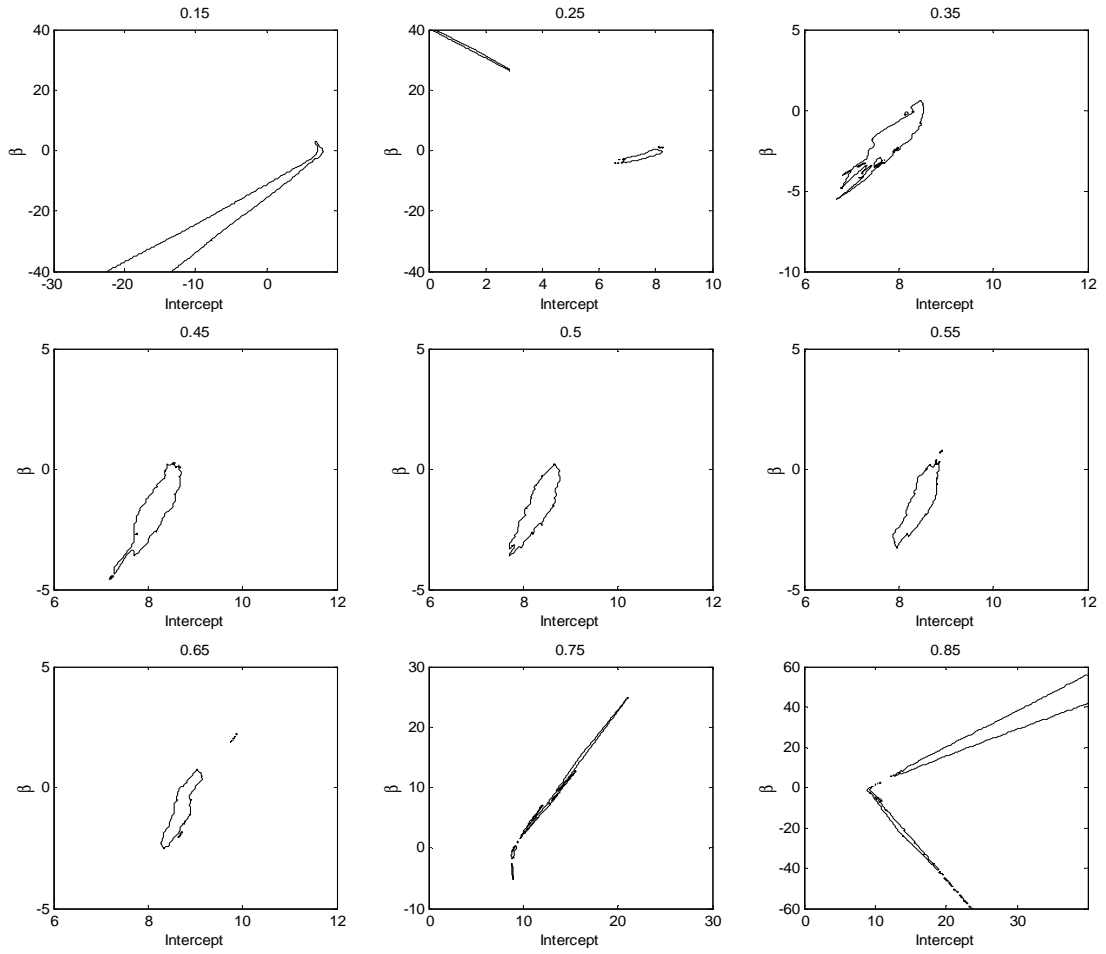
FIGURE 7. **Finite Sample Confidence Regions for Fish Example Treating Price as Endogenous.** This figure plots finite sample confidence regions from fish example without covariates treating price as endogenous. Values for the intercept, $\theta(\tau)_{[2]}$ are on the horizontal axis, and values for the slope parameter $\theta(\tau)_{[1]}$ are on the vertical axis.

36

TABLE 3. Returns to Schooling

Estimation Method	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
A. Quantile Regression (No Instruments)			
Quantile Regression (Asymptotic)	(0.0715,0.0731)	(0.0642,0.0652)	(0.0637,0.0650)
Finite Sample (MCMC)	(0.0710,0.0740)	(0.0640,0.0660)	(0.0637,0.0656)
Finite Sample (Grid)	(0.0710,0.0740)	(0.0641,0.0659)	(0.0638,0.0655)
Finite Sample (Marginal)	(0.0706,0.0742)	(0.0638,0.0662)	(0.0634,0.0658)
B. IV Quantile Regression (Quarter of Birth Instruments)			
Inverse Quantile Regression (Asymptotic)	(0.0784,0.2064)	(0.0563,0.1708)	(0.0410,0.1093)
Finite Sample (MCMC)	(0.1151,0.1491)	(0.0378,0.1203)	(0.0595,0.0703)
Finite Sample (MCMC-2)	(0.0580,0.2864)	(0.0378,0.1203)	(0.0012,0.0751)
Finite Sample (Grid)	(0.059,0.197)	(0.041,0.119)	(0.021,0.073)
Finite Sample (Marginal)	(0.05,0.39)	(0.03,0.13)	(0.00,0.08)
C. Quantile Regression - State and Year of Birth Effects			
(No Instruments)			
Quantile Regression (Asymptotic)	(0.0666,0.0680)	(0.0615,0.0628)	(0.0614,0.0627)
Finite Sample (Marginal)	(0.0638,0.0710)	(0.0594,0.0650)	(0.0590,0.0654)
D. IV Quantile Regression - State and Year of Birth Effects			
(Quarter of Birth Instruments)			
Inverse Quantile Regression (Asymptotic)	(0.0890,0.2057)	(0.0661,0.1459)	(0.0625,0.1368)
Finite Sample (Marginal)	(-0.24,1]	[-1,1]	[-1,0.35]

Note: 95% level confidence interval estimates for Returns to Schooling example. Panel A reports results from model which treats schooling as exogenous, and Panel B reports results from model which treats schooling as endogenous and uses quarter of birth dummies as instruments for schooling. Panels C and D are as A and B but include a set of 51 state of birth dummy variables and a set of 9 year of birth dummy variables. The first row in each panel reports the interval estimated using the asymptotic approximation, and the remaining rows report estimates of the finite sample interval constructed through various methods.
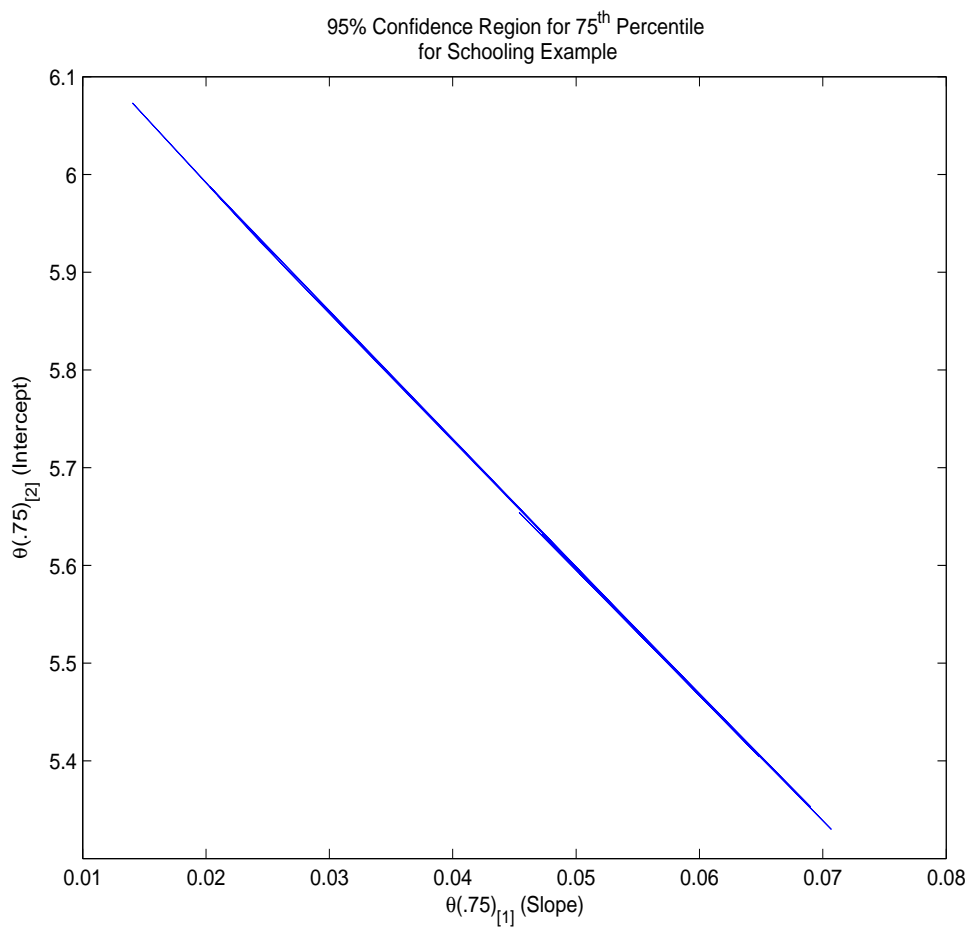
FIGURE 8. **Confidence Region for 75$^{\text{th}}$ Percentile in Schooling Example.** This figure plots the finite sample confidence regions from the schooling example in the model without covariates treating schooling as endogenous. Values for the intercept, $\theta(.75)_{[2]}$ are on the vertical axis, and values for the slope parameter $\theta(.75)_{[1]}$ are on the horizontal axis.