



## Finite sample inference for quantile regression models

Victor Chernozhukov<sup>a,\*</sup>, Christian Hansen<sup>b</sup>, Michael Jansson<sup>c</sup>

<sup>a</sup> Massachusetts Institute of Technology, Department of Economics, United States

<sup>b</sup> University of Chicago, Graduate School of Business, United States

<sup>c</sup> University of California - Berkeley, Department of Economics, United States

### ARTICLE INFO

#### Article history:

Available online 14 January 2009

#### JEL classification:

C12

C14

#### Keywords:

Extremal quantile regression  
Instrumental quantile regression  
Partial identification  
Weak identification

### 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 solve and provide valid finite sample inference for linear and nonlinear quantile models with endogenous or exogenous covariates. The confidence regions can be computed using Markov Chain Monte Carlo (MCMC) methods. We illustrate the finite sample procedure through two empirical examples: estimating a heterogeneous demand elasticity and estimating heterogeneous returns to schooling. We find pronounced differences between asymptotic and finite sample confidence regions in cases where the usual asymptotics are suspect.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

Quantile regression (QR) methods, initiated largely in the seminal work of Koenker and Bassett (1978), provide useful tools for examining the effects of covariates on an outcome variable of interest. Perhaps the most appealing feature of QR methods is that they allow 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 Koenker and Bassett (1978), Portnoy (1991), Buchinsky (1994), and Chamberlain (1994), among others. Koenker (2005) provides an excellent introduction to QR methods.

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 approach makes 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). We extend the results from the unconditional case to the estimation of regression 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 and related “pivotal” resampling methods, see e.g. Gutenbrunner and Jurečková (1992) and Parzen et al. (1994), but, in sharp contrast to these approaches, it does not require asymptotics or homoskedasticity (in the case of the rank score methods) 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, requiring only weak independence assumptions on the sampling mechanism and continuity of quantile functions in the probability index. In

\* Corresponding address: Massachusetts Institute of Technology, Department of Economics, 50 Memorial Drive, Cambridge, MA 02142, United States. Tel.: +1 617 253 4767.

E-mail address: [vchern@mit.edu](mailto:vchern@mit.edu) (V. Chernozhukov).

endogenous settings, the finite sample approach will remain valid in cases of weak identification or set identification (e.g. as in Haile and Tamer (2003) and Chernozhukov et al. (2007a)). 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 methods 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.

We 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 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

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–A4 corresponds to the conventional quantile regression model with exogenous covariates, see Koenker (2005), where  $Y$  is the dependent variable,  $D$  is the regressor, and  $q(d, \tau)$  is the  $\tau$ -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 Bhattacharya (1963), Doksum (1974), Hogg (1975), Koenker and Bassett (1978) and Matzkin (2003), as discussed in more detail in Koenker (2005). 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,  $Y = D'\theta(U)$ , were developed in Koenker and Bassett (1978), and estimation and inference results have been extended in a number of useful directions by subsequent authors. 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–A4 provides a generalization of the conventional quantile model that allows for endogeneity. See Chernozhukov and Hansen (2001, 2005, 2006, 2008) for discussion of the model as well as for semi-parametric estimation and inference theory under strong and weak identification. See Chernozhukov et al. (2007b) for a nonparametric analysis of this model and Chesher (2003) for a related nonseparable model. The model A1–A4 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 provides the basis for the finite sample inference results that follow.

**Proposition 1 (Main Statistical Implication).** Suppose A1–A3 hold, then

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

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

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

Eq. (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 (2005).

Eq. (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( $\tau$ ) random variable regardless of the sample size. This random variable depends only on  $\tau$ , 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( $\tau$ ) 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  $(Y_i, D_i, Z_i, i = 1, \dots, n)$  on probability space  $(\Omega, \mathcal{F}, P)$  (possibly dependent on the sample size), such that A1–A4 hold for each  $i = 1, \dots, n$ , and the following additional conditions hold:

**A5 (Parameterization):** There exists  $\theta_0 \in \Theta_n \subset \mathbb{R}^{K_n}$  such that  $q(D, \tau) = q(D, \theta_0, \tau)$ , with equality holding up to a numerically negligible error, where the function  $q(D, \theta, \tau)$  is known, but  $\theta_0$  is not.

**A6 (Conditionally Independent Sampling):**  $(U_1, \dots, U_n)$  are i.i.d. Uniform(0, 1), conditional on  $(Z_1, \dots, Z_n)$ .

We will use the letter  $\mathcal{P}$  to denote the collection of all probability laws  $P$  on the measure space  $(\Omega, \mathcal{F})$  that satisfy conditions A1–A6.

Conditions A5–A6 restrict the model A1–A4 sufficiently to allow finite sample inference. A5 requires that the  $\tau$ -quantile function  $q(d, \tau)$  can be approximated by a finite-dimensional model  $q(d, \tau, \theta_0)$  (where  $\theta_0$  may vary with  $\tau$ ), up to a negligible numerical error. The finite sample inference results of this paper in principle apply to any parameter space, including a function space. However, from a practical point of view, we should require that each element of the function space can be approximated in a suitable norm by a finite-dimensional model, and the approximation error can be made arbitrarily small. 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. Of course, the usual parametric assumptions are permitted. A5 allows the model to depend on the sample size 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  $K_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Condition A6 is obviously satisfied if  $(Y_i, X_i, Z_i, i = 1, \dots, n)$  are i.i.d., but in principle should allow for some dynamics, e.g. of the kinds considered in Koener and Xiao (2004a,b).

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 Eq. (2.1) in Proposition 1 justifies the following generalized method-of-moments (GMM) function for estimating  $\theta_0$ :

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

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  $Z$ , and  $W_n$  is a positive semi-definite weight matrix which is fixed conditional on  $Z_1, \dots, Z_n$ . One would typically choose  $g(Z)$  such that  $\dim(g(Z)) \geq \dim(\theta_0)$ , though this is not required for validity of the approach. 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  $Z_1, \dots, Z_n$ . Since this conditional variance does not depend on  $\theta_0$ , the GMM function with  $W_n$  defined above also corresponds to the continuous-updating estimator of Hansen et al. (1996). In all examples in this paper, we use the identity function for  $g(\cdot)$ .

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:

$L_n(\theta_0) \stackrel{d}{=} \mathcal{L}_n$ , conditional on  $(Z_1, \dots, Z_n)$ , where

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

and  $(B_1, \dots, B_n)$  are i.i.d. Bernoulli random variables with  $EB_i = \tau$ , which are independent of  $(Z_1, \dots, Z_n)$ .

**Proof.** Implication 2 of Proposition 1 and A6 imply the result.  $\square$

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

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  $\alpha$ -quantile of  $\mathcal{L}_n$ . By inverting this test-statistic, one obtains confidence regions for  $\theta_0$ .

Let  $CR(\alpha)$  be the  $c_n(\alpha)$ -level set of the function  $L_n(\theta)$ :  $CR(\alpha) \equiv \{\theta : L_n(\theta) \leq c_n(\alpha)\}$ . It follows immediately from the previous results that  $CR(\alpha)$  is a valid  $\alpha$ -level confidence region for  $\theta_0$ . This result is stated formally in Proposition 3.

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

**Proof.**  $\theta_0 \in CR(\alpha)$  is equivalent to  $\{L_n(\theta_0) \leq c_n(\alpha)\}$  and  $\Pr_P\{L_n(\theta_0) \leq c_n(\alpha)\} \geq \alpha$ , by the definition of  $c_n(\alpha) := \inf\{l : P\{\mathcal{L}_n \leq l\} \geq \alpha\}$  and  $L_n(\theta_0) \stackrel{d}{=} \mathcal{L}_n$ .  $\square$

Proposition 3 demonstrates how one may obtain valid finite sample confidence regions and tests for the parameter vector  $\theta$  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, see Gutenbrunner and Jurečková (1992) and Gutenbrunner et al. (1993), and the pivotal bootstrap method of Parzen et al. (1994). In contrast to the pivotal bootstrap and the rank-score method, the finite sample approach does not rely on asymptotics for its validity and is valid in finite samples. Moreover, the rank-score method relies on a homoskedasticity assumption, while the finite sample approach does not.

It is worth emphasizing the distinction between the finite sample approach and some other inferential procedures. The pivotal bootstrap produces a “fiducial” distribution, whose support is a finite set of points that generically does not contain the true parameter  $\theta_0$ . Therefore, the pivotal bootstrap does not possess formal finite sample validity. Furthermore, Parzen et al. (1994) establish its asymptotic validity under conditions of strong

identifiability and asymptotic normality for the QR estimator. We conjecture that the asymptotic validity of the pivotal bootstrap does not hold in more nonstandard settings such as cases with weak instruments, under set identification, or for extreme quantiles. In contrast, the finite sample approach is shown to work under all such conditions. The finite sample method should not be confused with the Gibbs bootstrap proposed in He and Hu (2002), which may be viewed as a computationally attractive alternative to Parzen et al. (1994). The method is also different from specifying the finite sample density of quantile regression as in Koenker and Bassett (1978). The finite sample density of QR is not pivotal and it cannot be used for finite sample inference unless the nuisance parameters (the conditional density of the outcome given the regressors) are specified.

Finally, we note that, while 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  $\theta$ , 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  $\theta$  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 nonlinear, nonseparable model with heterogeneous effects that allows for endogeneity.

The approach is valid without imposing distributional assumptions and allows for general forms of heteroskedasticity and some 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 for example in Chernozhukov and Hansen (2006).

In addition to these features, the finite sample inference procedure will remain valid in situations where the parameters of the model are only set-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  $\theta_0$  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 trivially that it remains valid under the asymptotic formalization of “weak instruments”, as defined for example 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  $K_n/n \rightarrow 0$ ,

$K_n \rightarrow \infty, n \rightarrow \infty$ . 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  $K_n^2/n \rightarrow 0, K_n \rightarrow \infty, n \rightarrow \infty$ .

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. One alternative to using the conventional asymptotic approximation for extremal quantiles is to pursue an approach explicitly aimed at performing inference for extremal quantiles, for example as in Chernozhukov (2005). 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 the inference approaches conventional GMM based joint inference, as for example in Pakes and Pollard (1989), Abadie (1995) and Chernozhukov et al. (2003).

Finally, it is important to note that inference is simultaneous on all components of  $\theta$  and that for joint inference the approach is not conservative. Inference about subcomponents of  $\theta$  may be made by projections, as illustrated in the previous section, and may be conservative.

#### 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. One possible approach is to simply use a naive grid-search, but as the dimension of  $\theta$  increases, this approach becomes intractable. To help alleviate this problem, we explore the use of MCMC methods. MCMC methods seem attractive in this setting because they generate 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 in some cases.

##### 2.5.1. Computation of the critical value

The computation of the critical value  $c_n(\alpha)$  may proceed by simulating the distribution  $\mathcal{L}_n$ . We outline a simulation routine below.

**Algorithm 1** (Computation of  $c_n(\alpha)$ ). Given  $(Z_i, i = 1, \dots, n)$ , for  $j = 1, \dots, J$ : 1. Draw  $(U_{i,j}, i \leq n)$  as i.i.d. 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  $\alpha$ -quantile of the sample  $(\mathcal{L}_{n,j}, j = 1, \dots, J)$ , 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 nonsmooth, nonconvex function. For even moderately 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. Of course, other MCMC algorithms or stochastic search methods could also be employed. 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 methods 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 et al. (2003), except that we use it here to get level sets of the objective function rather than pseudo-posterior means and quantiles. A related suggestion would be to use  $f(\theta) \propto \exp(-\max[L_n(\theta), c_n(\alpha)])$  as a quasi-posterior density. This choice of  $f(\theta)$  samples the confidence region uniformly.

Given  $f$ , the basic random walk MCMC algorithm is implemented as follows:

**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 the starting point for each repetition.

At each step, the MCMC algorithm considers two potential values for  $\theta$  and obtains the corresponding values of the objective function. Step 3 above differs from a conventional random walk MCMC algorithm 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  $h(\cdot)$ . 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  $h(\cdot)$  and the starting value.<sup>1</sup> 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  $\{\theta^{(1)}, \dots, \theta^{(k)}\}$  and, as a by-product, a set of values for the objective function  $\{L_n(\theta^{(1)}), \dots, L_n(\theta^{(k)})\}$ . 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  $\theta$  where the value of  $L_n(\theta) \leq c_n(\alpha)$ :  $\hat{CR}(\alpha) = \{\theta^{(i)} : L_n(\theta^{(i)}) \leq c(\alpha)\}$ .

### 2.5.3. Computation of confidence bounds for individual regression parameters

The MCMC approach outlined above may be used to estimate joint confidence regions. 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  $\theta$ , 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  $\theta$  with  $\theta_{[1]} = \theta_{[1]}^*$  that satisfies  $L_n(\theta) \leq c_n(\alpha)$ . Since only one such value of  $\theta$  is required to place  $\theta_{[1]}^*$  in the confidence bound, we may restrict consideration to  $\theta^*$ , 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  $\theta_{[1]}$ ,  $\{\theta_{[1]}^j, j = 1, \dots, J\}$ . 2. For  $j = 1, \dots, J$ , find  $\theta_{[-1]}^j = \arg \inf_{\theta_{[-1]}} L_n((\theta_{[1]}^j, \theta_{[-1]}^j)')$ . 3. Calculate the confidence region for  $\theta_{[1]}$  as  $\{\theta_{[1]}^j : L_n((\theta_{[1]}^j, \theta_{[-1]}^j)' ) \leq c_n(\alpha)\}$ .

In addition to its 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. 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 two simple case studies.<sup>2</sup> 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 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 methods, 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. 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

<sup>1</sup> In our applications, we use estimates of  $\theta$  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 (2006) in endogenous cases as starting values and transition densities.

<sup>2</sup> Simulation results are available in a previous working paper version of this paper, Chernozhukov et al. (2006). In the simulations, we find that tests about the entire parameter vector based on the finite sample method have the correct size and that tests about individual parameters based on the finite sample method have correct size in the sense that the size is less than the nominal level but may be conservative. In both cases, conventional asymptotic procedures have large size distortions in situations where the asymptotic approximations may be suspect. The results also suggest the finite sample procedure has nontrivial power.

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 nonadditive 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.<sup>3</sup> 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 (2006) 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). We use the Hall–Sheather bandwidth choice suggested by Koenker (2005) to implement the asymptotic standard errors.

Estimation results are presented in Table 1. Panel A of Table 1 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, and the marginal procedure in Panels A and B.<sup>4</sup> 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 basic results remain unchanged. Also, all three computational methods for obtaining the finite sample confidence bounds give similar answers in the model with only an intercept and price. 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 that the three approaches to constructing the finite sample interval in general give similar results in this case. This finding is graphically illustrated in Fig. 1 for instrumental variables estimates at the median. 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.<sup>5</sup>

### 3.2. Returns to schooling

As our final example, we consider estimation of a simple return 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 US 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  $\tau$ . Since we believe that years of schooling may be jointly determined with unobserved ability, we use quarter of

<sup>5</sup> Further evidence provided in Chernozhukov et al. (2006) suggests that the problem at  $\tau = .25$  is due to the confidence set's being disconnected. The simple MCMC algorithm explores one region of the confidence set but fails to jump to the other region. This problem could likely be remedied by employing a more sophisticated search algorithm.

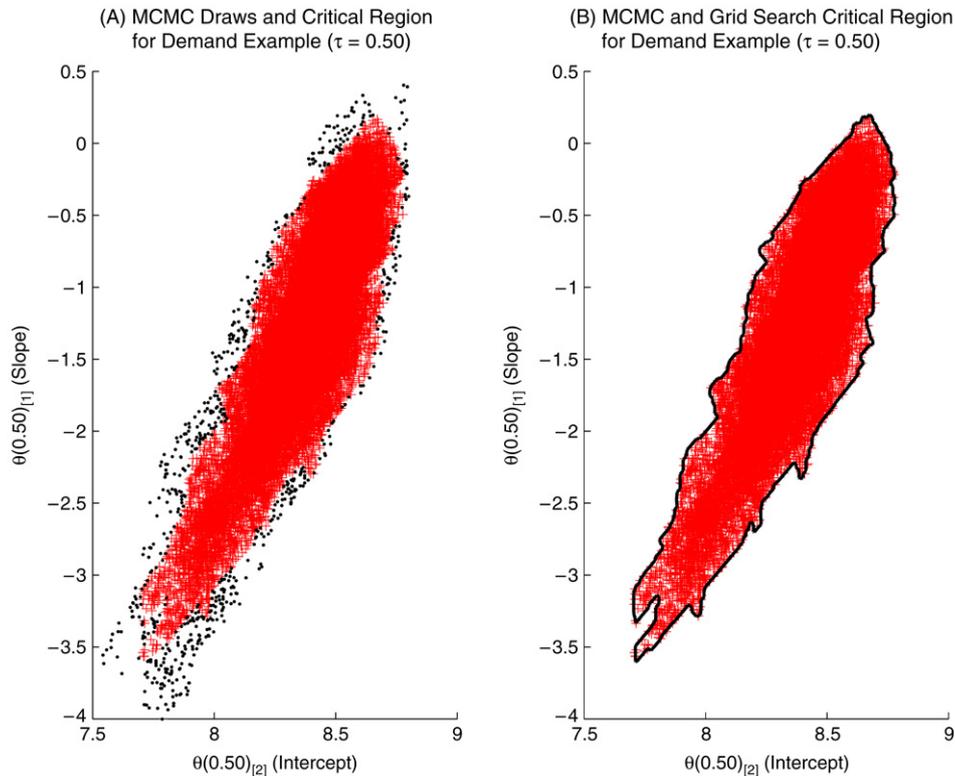
<sup>3</sup> More detailed arguments may be found in Graddy (1995).

<sup>4</sup> Details of the computation are described in Chernozhukov et al. (2006).

**Table 1**  
95% level confidence interval estimates for demand for fish example.

Estimation method	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
<b>Panel A. Quantile regression (No instruments)</b>			
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>Panel B. IV Quantile Regression (Stormy, Mixed as Instruments)</b>			
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]
<b>Panel C. Quantile regression – Day effects (No Instruments)</b>			
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)
<b>Panel D. IV Quantile regression – Day effects (Stormy, Mixed as Instruments)</b>			
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: 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.



**Fig. 1.** Computation of a confidence region by MCMC. The left panel shows the MCMC draws with the darker points corresponding to draws that fall outside of the confidence region. The solid line in the right panel is the grid search confidence region which is plotted against the MCMC confidence region.

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  $\beta(U)$ .

We present estimation results in Table 2. Panel A of Table 2 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, and the marginal procedure in Panels A and B.<sup>6</sup> The modified MCMC procedure we employ is a 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

<sup>6</sup> Details of the computation are described in Chernozhukov et al. (2006).

**Table 2**  
95% level confidence interval estimates for the returns to schooling example.

Estimation method	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
Panel 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)
Panel 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)
Panel 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)
Panel 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: 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.

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 is 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. 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. The difficulty in this case is that the objective function has extremely sharp “line”-like contours.<sup>7</sup> 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 coarse grid, one may miss the confidence region entirely. The narrowness of the confidence set causes problems with MCMC by making transitions quite difficult. The MCMC-2 procedure alleviates the problems with the random walk MCMC somewhat by running multiple chains with different starting values. 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.

#### 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 solves are conditionally pivotal in finite samples. This conditional pivotal property allows the construction of exact finite sample joint confidence regions and of 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. Finally, we illustrate the finite sample procedure in two empirical examples: estimation of a demand curve in a small sample and estimation of the returns to schooling in a large sample.

#### Acknowledgements

The finite sample results of this paper were included in the April 17, 2003 version of the paper “An IV Model of

<sup>7</sup> The shape of the contours is illustrated graphically in Fig. 8 of Chernozhukov et al. (2006).

Quantile Treatment Effects” (<http://gsbwww.uchicago.edu/fac/christian.hansen/research/IQR-short.pdf>). As a separate project, the current paper was prepared for the Winter Meetings of the Econometric Society in San Diego, 2004. We thank Roger Koenker for constructive discussion of the paper at the Meetings that led to the development of the optimality results for the inferential statistics used in the paper. We also thank Andrew Chesher, Lars Hansen, Jim Heckman, Marcello Moreira, Rosa Matzkin, Jim Powell, Whitney Newey, and seminar participants at Northwestern University for helpful comments and discussion. This research was supported by the William S. Fishman Faculty Research Fund and the IBM Corporation Faculty Research Fund at the University of Chicago Graduate School of Business.

**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. First, 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 et al. (1996).) Second, 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  $(Y_i, D_i, Z_i, i = 1, \dots, n)$  is an i.i.d. sample from the model defined by A1–A6 and assume that the dimension  $K$  of  $\theta_0$  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–A6,  $P[Y - q(D, \theta_0, \tau) \leq 0 \mid Z] = \tau$ . Consider the problem of testing  $H_0 : \theta_0 = \theta_*$  vs.  $H_a : \theta_0 \neq \theta_*$ , where  $\theta_* \in \mathbb{R}^K$  is some constant.

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

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

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

Being based on  $(e_i, Z_i, i = 1, \dots, n)$ , any such  $L_n(\theta_*, g)$  is conditionally pivotal under the null. In addition, under the null,  $L_n(\theta_*, g) \rightarrow_d \frac{1}{2} \chi_{\dim(g)}^2$  for any  $g \in \mathcal{G}$ . Moreover, the class  $\{L_n(\theta_*, g) : g \in \mathcal{G}\}$  enjoys desirable large sample power properties under the following strong identification assumption in which  $\Theta_*$  denotes some open neighborhood of  $\theta_*$ .

**Assumption 3.** (a) The distribution of  $Z$  does not depend on  $\theta_0$ . (b) For every  $\theta \in \Theta_*$  (and for almost every  $Z$ ),

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

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

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

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

where  $\delta_S(b, g) = b'E[\dot{\tau}_*(Z)g(Z)']E[g(Z)g(Z)']^{-1}E[g(Z)\dot{\tau}_*(Z)']b$ . By a standard argument,  $\delta_S(b, g) \leq \delta_S(b, \dot{\tau}_*)$  for any  $g \in \mathcal{G}$ . As a consequence,  $L_n(\theta_*, \dot{\tau}_*)$  maximizes the local asymptotic power within the class  $\{L_n(\theta_*, g) : g \in \mathcal{G}\}$ . An even stronger optimality result is the following.

**Proposition 4.** Among tests based on  $(e_i, Z_i, i = 1, \dots, n)$ , 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  $Z = (Z_1, \dots, Z_n)$ ) log likelihood function is given by  $\ell_n(\theta \mid Z) = \sum_{i=1}^n \{\log[\tau(Z_i, \theta)]e_i + \log[1 - \tau(Z_i, \theta)](1 - e_i)\}$ . Assumption 3 implies that the following LAN expansion is valid under the null. For any  $b \in \mathbb{R}^K$ ,  $\ell_n(\theta_* + \frac{b}{\sqrt{n}}) - \ell_n(\theta_*) = b'g_n^* - \frac{1}{2}b'J_n^*b + o_p(1)$ , where  $\ell_n$  is the (unconditional) log likelihood function,  $g_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{\tau(1-\tau)} \dot{\tau}_*(Z_i)(e_i - \tau) \rightarrow_d \mathcal{N}(0, J^*)$ , and  $J_n^* = \frac{1}{n} \sum_{i=1}^n \frac{1}{\tau(1-\tau)} \dot{\tau}_*(Z_i)\dot{\tau}_*(Z_i)' \rightarrow_p J^* = \frac{1}{\tau(1-\tau)} E[\dot{\tau}_*(Z)\dot{\tau}_*(Z)']$ . Theorem 3 of Choi et al. (1996) now shows that  $L_n(\theta_*, \dot{\tau}_*) = \frac{1}{2}g_n^{*'}J_n^{*-1}g_n^*$  is the asymptotically UMP invariant test of  $H_0$ .  $\square$

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

$$Y = D'\theta_0 + \varepsilon, \quad (\text{A.4})$$

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

$$D = \Pi'Z + v, \quad (\text{A.5})$$

where  $(\varepsilon, v') \mid Z \sim \mathcal{N}(0, \Sigma)$  for some positive definite matrix  $\Sigma$ , 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) assumes that  $\dot{\tau}_*$  is known.)

**Comment A.1 (Estimation of the Optimal Instrument).** If one is interested in using optimal instruments in practice, they need to be estimated. The asymptotic properties of the test will hold with estimated optimal instruments using the full sample as long as a consistent estimate of the gradient function which  $\hat{\tau}_*(\cdot)$  belongs to a Donsker class  $\mathcal{T}$  of functions is used; see Andrews (1994) for a list of parametric or nonparametric methods for estimation of the gradient that will satisfy this condition. In addition, estimation of the gradient will not affect the validity of the finite sample inference provided sample splitting is used. By the latter, we mean that consistent estimation of  $\tau_*(\cdot)$  and finite sample inference are performed using different subsamples of sizes  $b$  and  $n - b$  of the full sample of size  $n$ . If an asymptotically negligible fraction of the sample is used for the estimation of the gradient, i.e.,  $b/n \rightarrow 0$ , the first-order efficiency of the test is unaffected. In the case of the sample splitting, we can drop the technical requirement that  $\mathcal{T}$  is Donsker.

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 show that 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.<sup>8</sup>

**Assumption 4.** (a) The distribution of  $Z$  does not depend on  $\theta_0$ . (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)  $\lim_{n \rightarrow \infty} E[R_n(Z, \theta, C)^2] = 0$  for every  $\theta$  and every  $C$ .

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

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

where  $\delta_W(\Delta_\theta, C, g) = \Delta'_\theta E[C' Z g(Z)'] E[g(Z) g(Z)']^{-1} E[g(Z) Z' C] \Delta_\theta$ . As in the strongly identified case, the limiting distribution of  $L_n(\theta_*, g)$  is  $\frac{1}{2}$  times a noncentral  $\chi_{\dim(g)}^2$  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  $b, \delta_n^*$ , and  $J_n^*$  of the latter proof replaced by  $\Delta_\theta$ ,

$$\delta_n(C) = C' \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{\tau(1-\tau)} Z_i (e_i - \tau) \right] \quad \text{and}$$

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

respectively. (In particular, the proof utilizes the fact that, if  $C$  is known, then the statistic  $\delta_n(C)$  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

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

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  $\Sigma_{vv}, L_n^*$  is a strictly increasing transformation of

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

<sup>8</sup> Assumption 4 is motivated by the Gaussian model (A.4)–(A.5). In that model, parts (b) and (d) of Assumption 4 hold (with  $C$  proportional to  $\sqrt{n} \Pi$ ) if part (c) does and  $\Pi$  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  $\Sigma_{vv} \otimes (n^{-1} \sum_{i=1}^n Z_i Z_i')^{-1}$ . In (A.8), the functional form of  $J(\cdot)$  is “natural” insofar as it corresponds to the weak instruments prior employed by Chamberlain and Imbens (2004). Moreover, following Andrews and Ploberger (1995), 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  $\Delta_\theta \sim \mathcal{N}[0, \kappa J_n(C)^{-1}]$ . In view of the foregoing, it follows that the statistic  $L_n^*$  enjoys weighted average power optimality properties of the Andrews and Ploberger (1995) variety. As discussed by Andrews and Ploberger (1995, p. 1384),  $L_n^*$  can therefore be interpreted as (being asymptotically equivalent to) a Bayesian posterior odds ratio. This statement is formalized in the following result.

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

$$\limsup_{n \rightarrow \infty} \iint \Pr(\text{reject } \theta^* \mid \theta_0, C) dK_C(\theta_0) dJ(C).$$

**References**

Abadie, A., 1995. Changes in Spanish labor income structure during the 1980s: A quantile regression approach, CEMFI Working Paper No. 9521.  
 Andrews, D.W.K., 1994. Empirical process methods in econometrics. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. Elsevier, North-Holland.  
 Andrews, D.W.K., Ploberger, W., 1995. Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. Annals of Statistics 23, 1609–1629.  
 Angrist, J.D., Krueger, A., 1991. Does compulsory schooling attendance affect schooling and earnings. Quarterly Journal of Economics 106, 979–1014.  
 Bhattacharya, P.K., 1963. On an analog of regression analysis. Annals of Mathematical Statistics 34, 1459–1473.  
 Buchinsky, M., 1994. Changes in US 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: Sims, C. (Ed.), Advances in Econometrics. Elsevier, New York.  
 Chamberlain, G., Imbens, G., 2004. Random effects estimators with many instrumental variables. Econometrica 72, 295–306.  
 Chernozhukov, V., 2005. Extremal quantile regression. The Annals of Statistics 33 (2), 806–839.  
 Chernozhukov, V., Hansen, C., 2001. An IV model of quantile treatment effects, MIT Department of Economics Working Paper No. 02–06. www.ssrn.com.  
 Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. Econometrica 73 (1), 245–262.  
 Chernozhukov, V., Hansen, C., 2006. Instrumental quantile regression inference for structural and treatment effect models. Journal of Econometrics 132 (2), 491–525.  
 Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: A robust inference approach. Journal of Econometrics 142 (1), 379–398.  
 Chernozhukov, V., Hansen, C., Jansson, M., 2006. Finite sample inference for quantile regression models, MIT Department of Economics Working Paper No. 06–03. Available at SSRN.  
 Chernozhukov, V., Hong, H., 2003. An MCMC approach to classical estimation. Journal of Econometrics 115, 293–346.  
 Chernozhukov, V., Hong, H., Tamer, E., 2007a. Parameter set inference in a class of econometric models. Econometrica 1243–1285.  
 Chernozhukov, V., Imbens, G.W., Newey, W.K., 2007b. Instrumental variable estimation of nonseparable models. Journal of Econometrics 139 (1), 4–14.  
 Chesher, A., 2003. Identification in nonseparable models. Econometrica 71, 1405–1441.  
 Choi, S., Hall, W.J., Schick, A., 1996. Asymptotically uniformly most powerful tests in parametric and semiparametric models. Annals of Statistics 24, 841–861.  
 Doksum, K., 1974. Empirical probability plots and statistical inference for nonlinear models in the two-sample case. Annals of Statistics 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., Jurečková, J., 1992. Regression rank scores and regression quantiles. Annals of Statistics 20 (1), 305–330.  
 Gutenbrunner, C., Jurečková, J., Koenker, R., Portnoy, S., 1993. Tests of linear hypotheses based on regression rank scores. Journal of Nonparametric Statistics 2 (4), 307–331.  
 Haile, P., Tamer, E., 2003. Inference with an incomplete model of English actions. Journal of Political Economy 111, 1–51.  
 Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14 (3), 262–280.

- He, X., Hu, F., 2002. Markov chain marginal bootstrap. *Journal of the American Statistical Association* 97 (459), 783–795.
- Hogg, R.V., 1975. Estimates of percentile regression lines using salary data. *Journal of the American Statistical Association* 70, 56–59.
- 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. *Annals of Statistics* 1, 799–821.
- Koenker, R., 2005. *Quantile Regression*. Cambridge University Press.
- Koenker, R., Bassett, G.S., 1978. Regression quantiles. *Econometrica* 46, 33–50.
- Koenker, R., Xiao, Z., 2004a. Quantile autoregression, Working Paper. Available at [www.econ.uiuc.edu](http://www.econ.uiuc.edu).
- Koenker, R., Xiao, Z., 2004b. Unit root quantile autoregression inference. *Journal of the American Statistical Association* 99 (467), 775–787.
- MacKinnon, W.J., 1964. Table for both the sign test and distribution-free confidence intervals of the median for sample sizes to 1000. *Journal of the American Statistical Association* 59, 935–956.
- Matzkin, R.L., 2003. Nonparametric estimation of nonadditive random functions. *Econometrica* 71 (5), 1339–1375.
- Muirhead, R.J., 1982. *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79 (1), 147–168.
- Pakes, A., Pollard, D., 1989. Simulation and asymptotics of optimization estimators. *Econometrica* 57, 1027–1057.
- Parzen, M.I., Wei, L.J., Ying, Z., 1994. A resampling method based on pivotal estimating functions. *Biometrika* 81, 341–350.
- Portnoy, S., 1985. Asymptotic behavior of  $M$  estimators of  $p$  regression parameters when  $p^2/n$  is large. II. Normal approximation. *Annals of Statistics* 13 (4), 1403–1417.
- Portnoy, S., 1991. Asymptotic behavior of regression quantiles in nonstationary, dependent cases. *Journal of Multivariate Analysis* 38 (1), 100–113.
- Robert, C.P., Casella, G., 1998. *Monte Carlo Statistical Methods*. Springer-Verlag.
- Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.
- Stock, J.H., Wright, J.H., 2000. GMM with weak identification. *Econometrica* 68, 1055–1096.
- Walsh, J.E., 1960. Nonparametric tests for median by interpolation from sign tests. *Annals of the Institute of Statistical Mathematics* 11, 183–188.