

# Instrumental Variables Estimation and Weak-Identification-Robust Inference Based on Conditional Quantile Restriction

Vadim Marmer and Shinichi Sakata

University of British Columbia

Department of Economics

997-1873 East Mall

Vancouver, BC V6T 1Z1, Canada

Email: ssakata@interchange.ubc.ca

February 20, 2009 01:04:44 (UTC)

**INCOMPLETE DRAFT. Not for general circulation,  
quotation, or citation. Comments and feedbacks would  
be greatly appreciated. Please contact the author for  
the latest version.**

Extending the  $\mathcal{L}_1$ -IV estimator proposed by Sakata (1997, 2007), we develop a method, named the  $\rho_\tau$ -IV estimator, to estimate structural equations based on the conditional quantile restriction imposed on the error terms. We study its asymptotic behavior and show how to make statistical inferences on the regression parameters. Given practical importance of weak identification, a highlight of the paper is proposal of a test robust to the weak identification. The statistics used in our method can be viewed as a natural counterpart of the Anderson and Rubin's (1949) statistic in the  $\rho_\tau$ -IV estimation. As is the case for the Anderson-Rubin test in the standard IV regression, our test can have a low power, when there many instruments excluded from the regression equation. We show how we could avoid such problem

by taking for instruments a small number of linear combinations of all available instruments that are suitably selected based on data.

## 1 Introduction

In this paper, we develop a method, named the  $\rho_\tau$ -IV estimator, to estimate structural equations based on the conditional quantile restriction imposed on the error terms, extending the  $\mathcal{L}_1$ -IV estimator proposed by Sakata (1997, 2007). We study its large sample behavior and show how to make statistical inferences on the regression parameters, which allows a researcher to make statistical inferences on the regression parameters of interest. We pay particular attention to testing of hypotheses on the regression parameters under weak identification, as the weak identification is as an important possibility in the regression based on a conditional quantile restriction as that based on the conditional mean restriction. We propose a weak-identification-robust test that can be viewed as a natural counterpart of the Anderson and Rubin's (1949) statistic in  $\rho_\tau$ -IV estimation. As is the case for the Anderson-Rubin (AR) test in the standard IV regression, our test can have a low power, when many instruments are excluded from the regression equation. We show how we could avoid such problem by taking for instruments a small number of linear combinations of all available instruments that are suitably selected based on data.

The conventional instrumental variable (IV) estimator is typically based on the identification of the structural parameters through the conditional mean restriction that the mean of the structural error term conditional on a set of instrumental variables is zero. The conditional mean restriction is appealing, because, unlike the independence between the error term and the instruments, it does not impose restrictions on other features of the conditional distribution of the error term such as the variance of it.

Nevertheless, the conditional mean restriction is considered unsuitable in some applications. The conditional mean of a random variable critically depends on the tails of the conditional distribution of the variable. A small change in the tails can cause a large change in the conditional mean. In many applications, on the other hand, we know little about the part of the population distribution that correspond to the tails of the error distribution. This often makes it difficult to justify the conditional mean restriction.

The conditional mean restriction is not the only natural way to identify the parameters of structural equations. In many applications, the conditional mean restriction comes from informal intuition that the

“location” of the conditional distribution of the error term given a suitably chosen set of instruments should be constant. When we are faced by the above-mentioned concern about the conditional mean restriction, one would desire to capture the location of the conditional distribution of the error term by a measure that does not depend on tails. The conditional quantiles of the error term are examples of such location measures.

Sakata (1997, 2007) proposes identifying and estimating the regression parameters based on the conditional median restrictions. Chernozhukov and Hansen (2001, 2006) also consider identification of the regression parameters based the conditional quantile restrictions and propose an estimation method, taking an approach related to but different from Sakata’s. In the current paper, we extend the estimator of Sakata (1997, 2007) to propose an method called the  $\rho_\tau$ -IV method to estimate regression models with the conditional  $\tau$ -quantile restriction.

Being based on the same identification condition, our estimator is closely related to Chernozhukov and Hansen’s estimator. The computation burden of the two estimators are also comparable, as should be clear from the discussion in Section 3 of the current paper. A benefit of our approach is that the objective function to be maximized in the  $\rho_\tau$ -IV estimation takes a form similar to the “variance ratio” in the (normal) limited information maximum likelihood (LIML) estimation. This allows us to formulate a statistic analogous to the AR statistic, with which we can make weak-identification-robust inferences on the regression parameters of interest.

In the IV regression literature, many researchers have been paying attention to possible identification issues. Sargan (1983) points out that it is problematic that identification is nearly violated. The analysis of Phillips (1984, 1985) on the exact finite sample distribution of LIML clearly shows that lack of identifiability in structural equation estimation keeps the LIML estimator from consistently estimating the coefficient of the structural equation. Hillier (1990) also shows analogous results in considering the directional estimation of the coefficients of structural equations. Choi and Phillips (1992) further explores the behavior of the IV estimator under lack of identifiability.

When instrumental variables are poorly correlated with endogenous explanatory variables in linear regression, the distribution of the IV estimator is quite different from what the standard large sample theory suggests, as demonstrated by Nelson and Startz (1990b, 1990a) and Bound, Jaeger, and Baker (1995). Staiger and Stock (1997) propose an alternative way to approximate the distribution of the IV estimator with weak instruments. Stock and Wright (2000) then establish a way to approximate the distribution of

generalized-method-moments (GMM) estimators under weak identification.

Given the absence of convenient and reliable approximation to the distribution of the IV estimator with weak instruments, it is difficult to perform tests of hypotheses on regression parameters in the usual style (i.e., the  $t$ -test, the Wald test, etc.). On the other hand, the AR test originally proposed in Anderson and Rubin (1949) is not affected by weakness of instruments. For this reason, Staiger and Stock (1997) and Dufour (1997) recommend use of the AR test. The AR test even has nice power properties if the number of instruments is equal to the number of endogenous explanatory variables (Moreira 2001, Andrews, Moreira, and Stock 2004). The AR test, however, loses its power when the number of instruments is larger than the number of endogenous explanatory variables. To overcome this difficulty, Kleibergen (2002, 2005) and Moreira (2001) propose Lagrange multiplier tests. Also, Moreira (2003) introduces a way to robustify a wide range of tests to the problem of weak instruments.

The weak identification is as important a possibility in regression based on a conditional quantile restriction as in regression based on the conditional mean restriction. To this end, we propose a test that has an asymptotically correct size regardless of whether the identification is strong or not. The hypothesis we consider is that some regression parameters are equal to prespecified values. If we apply the  $\rho_\tau$ -IV method imposing the constraints of the null hypothesis, the objective function in the  $L_\tau$ -IV estimation maximized subject to the parameter constraints of the null hypothesis tends to be close to one under the null. The constraint maximum of the objective function is similar to the Anderson and Rubin (1949) statistic in the sense that it captures how much of the fitted structural error can be explained by the instruments. It, ranging between zero and one, is closed to one if the fitted structural residuals cannot be fitted by the instruments in the sample. Its value far from one is thus taken as an evidence against the null in our test. If the conditions in the null hypothesis include the coefficients of all regressors potentially weakly related to the instruments excluded from the regression function, then the proposed test involves no weak identification problem, so that our test is robust to weak-identification. As is the case for the AR test in the standard IV regression, our test can have a low power, when there many instruments excluded from the regression equation. We show how we could avoid such problem by taking for instruments a small number of linear combinations of all available instruments that are suitably selected based on data.

Our test is closely related to Chernozhukov and Hansen (2008). They formulate a test in a way convenient in the estimation framework of Chernozhukov and Hansen (2001, 2006), while we propose a test convenient

in the  $\mathcal{L}_1$ -IV estimation. Unlike their work, we consider an extra step to avoid losing power in the case with a highly overidentified model. Another work related to our test is Jun (2008). Jun formulates a test adapting the approach of Kleibergen (2005).

The rest of the paper is organized as follows. We first describe the basic setup and define the  $\rho_\tau$ -IV estimator in Section 2. Then, after briefly discussing the computation of the  $\rho_\tau$ -IV estimator in Section 3, we establish the consistency and asymptotic normality of the  $\rho_\tau$ -IV estimator and explains how to consistently estimate the asymptotic covariance matrix of  $\rho_\tau$ -IV estimator in Section 4. In Section 5, we develop a weak-identification-robust method to test hypotheses on the regression parameters and consider how to avoid losing power when a model is over-identified. Throughout the paper,  $|\cdot|$  denotes the Euclidean norm, and Limits are taken along the sequence of sample sizes growing to infinity, unless otherwise indicated.

## 2 $\rho_\tau$ -IV Estimator

ASSUMPTION 1: *Let  $(\Omega, \mathcal{F}, P)$  be a probability space. The data are a realization of an independently and identically distributed stochastic process  $\{X_t \equiv (y_t, Y_t', Z_t')' : \Omega \rightarrow \mathbb{R} \times \mathbb{R}^g \times \mathbb{R}^k\}_{t \in \mathbb{N}}$  such that  $E[|X_1|] < \infty$ , and for each  $c \in \mathbb{R}^{1+g+k} \setminus \{0\}$ ,  $P[c'X_t = 0] < 1$ .*

Partition  $Z_t$  as  $Z_t = (Z_{t,1}', Z_{t,2}')'$ , where  $Z_{t,1}$  is  $k_1 \times 1$ , and  $Z_{t,2}$  is  $k_2 \times 1$  (so that  $k_1 + k_2 = k$ ). The parameter of interest is the coefficients in regression of  $y_t$  on  $Y_t$  and  $Z_{t,1}$  described in the next assumption.

ASSUMPTION 2: *The subset  $B$  of  $\mathbb{R}^g$  is nonempty and compact. There exists a unique  $\theta_0 \in (\beta_0', \alpha_0')' \in B \times \mathbb{R}^{k_1}$  such that for each  $n \in \mathbb{N}$ , the conditional  $\tau$ -quantile of*

$$U_1 \equiv y_1 - Y_1' \beta_0 - Z_{1,1}' \alpha_0$$

*given  $Z_1$  is zero, where  $\tau$  is a known real constant in  $(0, 1)$ .*

If instead the conditional  $\tau$ -quantile of  $U_1$  given  $Y_1$  and  $Z_{1,1}$  is known to be zero,  $\beta_0$  and  $\alpha_0$  could be consistently estimated by the estimator of Koenker and Bassett (1978). In our current setup, Koenker and Bassett's estimator is inconsistent in general.

We here propose an estimator of the structural regression coefficients, following the approach described

in Section 11 of Sakata (2007). Define  $\rho_\tau : \mathbb{R} \rightarrow \mathbb{R}$  by

$$\rho_\tau(v) \equiv (\tau - 1(v < 0))v, \quad v \in \mathbb{R},$$

where  $1(A)$  is the indicator function that becomes one if and only if the condition  $A$  is true. Also define functions  $R : \mathbf{B} \times \mathbb{R}^{k_1} \times \mathbb{R}^k \rightarrow \mathbb{R}$  and  $Q : \mathbf{B} \times \mathbb{R}^{k_1} \rightarrow \mathbb{R}$  by

$$R(\beta, \alpha, \gamma) \equiv E[\rho_\tau(y_1 - Y_1'\beta - Z_{1,1}'\alpha - Z_1'\gamma)], \quad (\beta, \alpha, \gamma) \in \mathbf{B} \times \mathbb{R}^{k_1} \times \mathbb{R}^k$$

and

$$Q(\beta, \alpha) \equiv \frac{\inf_{\gamma \in \mathbb{R}^k} R(\beta, \alpha, \gamma)}{R(\beta, \gamma, 0)}, \quad (\beta, \alpha) \in \mathbf{B} \times \mathbb{R}^{k_1},$$

where  $R(\beta, \gamma, 0) > 0$  by the linear independence of the elements of  $X_t = (y_t, Y_t', Z_t)'$  required by Assumption 1. Because of the conditional  $\tau$ -quantile restriction imposed on  $U_1$  in Assumption 2, we have that  $Q(\beta_0, \alpha_0) = 1$ , so that for each  $(\beta, \alpha) \in \mathbf{B} \times \mathbb{R}^{k_1}$

$$0 \leq Q(\beta, \alpha) \leq Q(\beta_0, \alpha_0) = 1. \tag{1}$$

It follows that  $\theta_0 \in (\beta_0', \alpha_0)'$  is the maximizer of  $Q$  over  $\Theta \equiv \mathbf{B} \times \mathbb{R}^{k_1}$ .

Our estimator is the maximizer of the sample counterpart of  $Q$ , which is given by a sequence of random functions  $\{\hat{Q}_n : \mathbf{B} \times \mathbb{R}^{k_1} \times \Omega \rightarrow \mathbb{R}\}_{n \in \mathbb{N}}$  defined by

$$\hat{Q}_n(\beta, \alpha, \omega) \equiv \begin{cases} \frac{\inf_{\gamma \in \mathbb{R}^{k_2}} \hat{R}_n(\beta, \alpha, \gamma, \omega)}{\hat{R}_n(\beta, \alpha, 0, \omega)}, & \text{if } \inf_{(b, a) \in \mathbf{B} \times \mathbb{R}^{k_1}} \hat{R}_n(b, a, 0, \omega) > 0, \\ 1, & \text{otherwise,} \end{cases}$$

$$(\beta, \alpha) \in \mathbf{B} \times \mathbb{R}^{k_1}, \omega \in \Omega, n \in \mathbb{N},$$

where

$$\hat{R}_n(\beta, \alpha, \gamma, \omega) \equiv n^{-1} \sum_{t=1}^n \rho_\tau(y_t(\omega) - Y_t(\omega)'\beta - Z_{t,1}(\omega)'\alpha - Z_t(\omega)'\gamma),$$

$$(\beta, \alpha, \gamma) \in \mathbf{B} \times \mathbb{R}^{k_1} \times \mathbb{R}^k, \omega \in \Omega, n \in \mathbb{N}.$$

We now define our estimator.

**DEFINITION 1 (THE  $\rho_\tau$ -IV ESTIMATOR):** *Given Assumption 1, a sequence of random vectors  $\{\hat{\theta} \equiv (\hat{\beta}_n', \hat{\alpha}_n)'\} : \Omega \rightarrow \mathbf{B} \times \mathbb{R}^{k_1}\}_{n \in \mathbb{N}}$  is called the  $\rho_\tau$ -IV estimator if for each  $n \in \mathbb{N}$ ,  $\hat{Q}_n(\hat{\beta}_n, \hat{\alpha}, \cdot) = \sup_{(\beta, \alpha) \in \mathbf{B} \times \mathbb{R}^{k_1}} \hat{Q}_n(\beta, \alpha, \cdot)$ .*

For each  $(\beta, \alpha) \in \mathbf{B} \times \mathbb{R}^{k_1}$ , we have that

$$\begin{aligned} \inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(\beta, \alpha, \gamma, \cdot) &= \inf_{(\gamma_1, \gamma_2) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}} n^{-1} \sum_{t=1}^n \rho_\tau(y_t - Y_t' \beta - Z_{t,1}'(\alpha + \gamma_1) - Z_{t,2}' \gamma_2) \\ &= \inf_{(\gamma_1, \gamma_2) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}} n^{-1} \sum_{t=1}^n \rho_\tau(y_t - Y_t' \beta - Z_{t,1}' \gamma_1 - Z_{t,2}' \gamma_2) = \inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(\beta, 0, \gamma, \cdot). \end{aligned}$$

Given this fact, it holds that whenever  $\hat{R}_n(\beta, \alpha, 0, \cdot) > 0$  for every  $(\beta, \alpha) \in \mathbf{B} \times \mathbb{R}^{k_1}$ ,

$$\sup_{\alpha \in \mathbb{R}^{k_1}} \hat{Q}_n(\beta, \alpha, \cdot) = \sup_{\alpha \in \mathbb{R}^{k_1}} \frac{\inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(\beta, 0, \gamma, \cdot)}{\hat{R}_n(\beta, \alpha, 0, \cdot)} = \frac{\inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(\beta, 0, \gamma, \cdot)}{\inf_{\alpha \in \mathbb{R}^{k_1}} \hat{R}_n(\beta, \alpha, 0, \cdot)}, \quad \beta \in \mathbf{B}. \quad (2)$$

Because the numerator and denominator of the ratio on the right-hand side of (2) are continuous in  $\beta$ ,  $\sup_{\alpha \in \mathbb{R}^{k_1}} \hat{Q}_n(\beta, \alpha, \cdot)$  is continuous in  $\beta$  in all realizations, whenever  $\hat{R}_n(\beta, \alpha, 0, \cdot) > 0$  for every  $(\beta, \alpha) \in \mathbf{B} \times \mathbb{R}^{k_1}$ .

The continuity of  $\sup_{\alpha \in \mathbb{R}^{k_1}} \hat{Q}_n(\beta, \alpha, \cdot)$  in  $\beta$  is also satisfied when  $\hat{R}_n(\beta, \alpha, 0, \cdot)$  can touches zero, because  $\hat{Q}_n(\beta, \alpha, \cdot) = 1$  in such case. Thus, given the compactness of  $\mathbf{B}$ ,  $\rho_\tau$ -IV estimator  $\hat{\beta}_n$  of  $\beta_0$  exists by the standard result on the existence of extremum estimators such as Gallant and White (1988, Theorem 2.2).

Further,  $\rho_\tau$ -IV estimator is the solution of

$$\inf_{\alpha \in \mathbb{R}^{k_1}} \hat{R}_n(\hat{\beta}_n, \alpha, 0, \cdot) = \inf_{\alpha \in \mathbb{R}^{k_1}} \rho_\tau((y_t - Y_t' \hat{\beta}_n) - Z_{t,1}' \alpha).$$

That is, it is the Koenker and Bassett's (1978) quantile regression estimator taking  $(y_t - Y_t' \hat{\beta}_n)$  for the dependent variable and  $Z_{t,1}$  for the regressors, which surely exists.

**THEOREM 2.1:** *Given Assumption 1, the  $\rho_\tau$ -IV estimator exists.*

*Remark.* We could avoid the compactness requirement of  $\mathbf{B}$  by first defining the  $\rho_\tau$ -IV directional estimator, as Sakata (2007) does, and then deriving the slope estimator in Definition 1 from it. We, however, directly define the slope estimator by imposing compactness on  $\mathbf{B}$  for saving space in this paper.

### 3 Computation of the $\rho_\tau$ -IV Estimator

We could calculate the  $\rho_\tau$ -IV estimator, adapting the algorithm described in Sakata (2007) for the case  $\tau = 0.5$  in the straightforward manner. Sakata's algorithm is, however, slow if  $k_1$  is large, because it uses a global search algorithm to minimizes  $\hat{Q}_n$  over  $\mathbf{B} \times \mathbb{R}^{k_1}$ .

Given a  $\beta$ , however, the ratio on the right-hand side of (2) can be quickly calculated, because the minimization problems appearing in both the numerator and denominator of the ratio can be rewritten as

linear programming problems, as Koenker and Bassett explains. Thus, the  $\rho_\tau$ -IV estimator can be calculated by maximizing the ratio in terms of  $\beta$  over  $\mathbf{B}$ . Because the ratio may have local maximum, it is advisable to use a global search algorithm such as the simulated annealing algorithm in calculating  $\hat{\beta}_n$ , while  $\hat{\alpha}_n$  is the solution of the minimization problem in the denominator calculated with  $\hat{\beta}_n$ .

## 4 Large Sample Properties of the $\rho_\tau$ -IV Estimator

In investigating the consistency of the  $\rho_\tau$ -IV Estimator, it is convenient to the population counterpart of (2), i.e.,

$$\sup_{\alpha \in \mathbb{R}^{k_1}} Q(\beta, \alpha) = \sup_{\alpha \in \mathbb{R}^{k_1}} \frac{\inf_{\gamma \in \mathbb{R}^k} R(\beta, 0, \gamma)}{R(\beta, \alpha, 0)} = \frac{\inf_{\gamma \in \mathbb{R}^k} R(\beta, 0, \gamma)}{\inf_{\alpha \in \mathbb{R}^{k_1}} R(\beta, \alpha, 0)}, \quad \beta \in \mathbf{B}. \quad (3)$$

By Assumption 2,  $\beta \mapsto \sup_{\alpha \in \mathbb{R}^{k_1}} Q(\beta, \alpha) : \mathbf{B} \rightarrow \mathbb{R}$  is a continuous function uniquely maximized at  $\beta_0$ . We can also show that  $\{\inf_{\alpha \in \mathbb{R}^{k_1}} \hat{Q}_n(\beta, \alpha, \cdot)\}_{n \in \mathbb{N}}$  converges to  $\inf_{\alpha \in \mathbb{R}^{k_1}} Q(\beta, \alpha)$  uniformly in  $\beta \in \mathbf{B}$  a.s.- $P$  (Lemma A.3). By the standard result on consistency of extremum estimators (e.g., Pötscher and Prucha 1991, Lemma 4.2), we can establish the consistency of  $\{\hat{\beta}_n\}_{n \in \mathbb{N}}$  for  $\beta_0$ .

The estimator  $\hat{\alpha}_n$ , on the other hand, minimizes  $\hat{R}_n(\hat{\beta}_n, \alpha, 0, \cdot)$  with respect to  $\alpha$  over  $\mathbb{R}^{k_1}$ . Given the strong consistency  $\{\hat{\beta}_n\}$  for  $\beta_0$ , we can verify the a.s.- $P$  convergence of  $\hat{R}_n(\hat{\beta}_n, \alpha, 0, \cdot)$  to  $R(\beta_0, \alpha, 0)$  for each  $\alpha \in \mathbb{R}^{k_1}$  and utilize the convexity of  $\hat{R}_n(\hat{\beta}_n, \alpha, 0, \cdot)$  in  $\alpha$  to establish the strong consistency of  $\hat{\alpha}_n$  for  $\alpha_0$ .

**THEOREM 4.1:** *Under Assumptions 1 and 2,  $\{\hat{\theta}_n = (\hat{\beta}'_n, \hat{\alpha}'_n)'\}_{n \in \mathbb{N}}$  converges to  $\theta_0 = (\beta'_0, \alpha'_0)'$ .*

In establishing the asymptotic normality of the  $\rho_\tau$ -IV estimator, we impose the additional conditions stated in the next theorem.

**ASSUMPTION 3:** (a) *The minimizer of  $R(\beta_0, \alpha_0, \cdot) : \mathbb{R}^k \rightarrow \mathbb{R}$  over  $\mathbb{R}$  is unique (hence, it is uniquely minimized at the origin).*

(b) *The vector  $\beta_0$  is interior to  $\mathbf{B}$ . Also, a neighborhood  $B_0 \subset \mathbf{B}$  of  $\beta_0$ , a neighborhood  $A_0 \subset \mathbb{R}^{k_1}$  of  $\alpha_0$ , and a neighborhood  $\Gamma_{2,0} \subset \mathbb{R}^{k_2}$  of the origin satisfy the following conditions:*

(i) *The conditional distribution  $y_1$  given  $Y_1$  and  $Z_1$  has a continuous probability density function (pdf)  $f(\cdot | Y_1, Z_1)$  at  $Y'_1 \beta + Z'_{1,1} \alpha + Z'_{1,2} \gamma_2$  for each  $(\beta, \alpha, \gamma) \in B_0 \times A_0 \times \Gamma_{2,0}$  a.s.- $P$ .*

- (ii) There exists a random variable  $D : \Omega \rightarrow \mathbb{R}$  with a finite absolute moment such that for each  $\beta \in B_0$ , each  $\alpha \in A_0$ , and each  $\gamma_2 \in \Gamma_{2,0}$ ,

$$f(Y_1' \beta + Z_{1,1}' \alpha + Z_{1,2}' \gamma_2 | Y_1, Z_1)(|Y_1|^2 + |Z_1|^2) < D. \quad (4)$$

- (c) Let  $J$  be the Hessian of  $R$  at  $(\beta'_0, \alpha'_0, 0'_{k_2 \times 1})'$  and partition it as

$$J \equiv \begin{pmatrix} J_{\beta\beta} & J_{\beta\alpha} & J_{\beta\gamma} \\ J_{\alpha\beta} & J_{\alpha\alpha} & J_{\alpha\gamma} \\ J_{\gamma\beta} & J_{\gamma\alpha} & J_{\gamma\gamma} \end{pmatrix},$$

where  $J_{\beta\beta}$  is  $g \times g$ ,  $J_{\alpha\alpha}$  is  $k_1 \times k_1$ , and  $J_{\gamma\gamma}$  is  $k \times k$ . Then the matrix

$$J_{\theta\theta} \equiv \begin{pmatrix} J_{\beta\beta} & J_{\beta\alpha} \\ J_{\alpha\beta} & J_{\alpha\alpha} \end{pmatrix}$$

is positive definite, and  $J_{\gamma\theta} \equiv (J_{\gamma\beta}, J_{\gamma\alpha})$  is of full column rank.

- (d)  $E[|Y_1|^2 + |Z_1|^2] < \infty$ .

Assumption 3(b) ensures the twice continuous differentiability of  $R$  in a neighborhood of  $(\beta_0, \alpha_0, 0)$  in  $\mathbb{R}^g \times \mathbb{R}^{k_1} \times \mathbb{R}^k$ , which then implies the twice continuous differentiability of  $Q$  in a neighborhood of  $\theta_0 = (\beta'_0, \alpha'_0)'$ . The first condition in Assumption 3(c) ensures that the Hessian of  $R(\beta_0, \alpha_0, \cdot) : \mathbb{R}^k \rightarrow \mathbb{R}$  at its minimum is negative definite. Under these conditions, the Hessian of  $(\beta, \alpha) \mapsto \log Q(\beta, \alpha) : \mathbf{B} \times \mathbb{R}^{k_1} \rightarrow \mathbb{R}$  at  $(\beta'_0, \alpha'_0)'$  is guaranteed to be positive definite, being equal to  $-K$ , where  $K \equiv R(\beta_0, \alpha_0, 0)^{-1} J_{\theta\gamma} J_{\gamma\gamma}^{-1} J_{\gamma\theta}$ . The full column rankness of  $J_{\gamma\beta}$  means that within a neighborhood of  $(\beta'_0, \alpha'_0)'$ , moving  $(\beta', \alpha')'$  away from  $(\beta'_0, \alpha'_0)'$  causes the gradient of  $R(\beta, \alpha, \cdot) : \mathbb{R}^k \rightarrow \mathbb{R}$  to be bounded away from zero uniformly in all directions, so that we can choose  $\gamma$  to make  $R(\beta, \alpha, \gamma)$  smaller than  $R(\beta, \alpha, 0)$ , once  $(\beta', \alpha')'$  deviates from  $(\beta'_0, \alpha'_0)'$ . Assumption 3(d) ensures that the Lindeberg-Levy Central Limit Theorem (Rao 1973, p. 127) applies the generalized score of the  $\rho_\tau$ -IV estimator. The moment requirements in Assumptions 3(b,d) are mild. If  $f(y_1 - Y_1' \beta - Z_{1,1}' \alpha - Z_{1,1}' \gamma | Y_1, Z_1)$  is bounded, they merely require that each element of  $Y_1$  and  $Z_1$  has a finite second moment, while the asymptotic normality of the conventional IV estimator is typically established under the assumption that the fourth moments of the dependent variable, the regressors, and the instruments are finite.

LEMMA 4.2: *Suppose that Assumption 1–3 hold. Let  $\{b_n\}_{n \in \mathbb{N}}$  and  $\{a_n\}_{n \in \mathbb{N}}$  be sequences of  $\mathbf{B}$ - and  $\mathbb{R}^{k_1}$ -valued random vectors, respectively. Then there exists a sequence of  $k \times 1$  random vectors  $c_n$  such that for each  $n \in \mathbb{N}$ ,  $\hat{R}_n(b_n, a_n, c_n, \cdot) = \inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(b_n, a_n, \gamma, \cdot)$ . If, in addition, Assumptions 3 hold, and  $b_n \rightarrow \beta_0$  and  $a_n \rightarrow \alpha_0$  in probability- $P$ , then*

$$n^{1/2}c_n = Cn^{1/2} \begin{pmatrix} b_n - \beta_0 \\ a_n - \alpha_0 \end{pmatrix} + J_{\gamma\gamma}^{-1}n^{-1/2} \sum_{t=1}^n (\tau - 1(U_t < 0)) Z_t + o_P(n^{1/2}|b_n - \beta_0| + n^{-1/2}|a_n - \alpha_0| + 1),$$

and  $C \equiv -J_{\gamma\gamma}^{-1}J_{\gamma\theta}$ .

Using this lemma, we can now approximate  $\log \hat{Q}_n$ .

LEMMA 4.3: *Suppose that Assumptions 1–3 hold and let  $\{b_n\}_{n \in \mathbb{N}}$  and  $\{a_n\}_{n \in \mathbb{N}}$  be sequences of  $\mathbf{B}$ - and  $\mathbb{R}^{k_1}$ -valued random vectors, respectively, that converge to  $\beta_0$  and  $\alpha_0$ . Write  $\tilde{\theta}_n \equiv (b'_n, a'_n)'$ . Then*

$$\begin{aligned} n \log \hat{Q}_n(b_n, a_n, \cdot) &= -\frac{1}{2R(\beta_0, \alpha_0, 0)} \left( n^{-1/2} \sum_{t=1}^n (\pi - 1(U_t < 0)) Z_t \right)' J_{\gamma\gamma}^{-1} \left( n^{-1/2} \sum_{t=1}^n (\pi - 1(U_t < 0)) Z_t \right) \\ &\quad - \frac{1}{R(\beta_0, \alpha_0, 0)} n^{-1/2} \sum_{t=1}^n (\pi - 1(U_t < 0)) Z_t' C n^{1/2} (\tilde{\theta}_n - \theta_0) - \frac{1}{2} n^{1/2} (\tilde{\theta}_n - \theta_0)' K n^{1/2} (\tilde{\theta}_n - \theta_0) \\ &\quad + o_P(n^{1/2}|b_n - \beta_0| + n|b_n - \beta_0|^2 + 1). \end{aligned} \tag{5}$$

Given this lemma, it is natural to expect that the minimizer of the the first three terms on the right-hand side of (5) approximates  $\hat{\theta} = (\hat{\beta}'_n, \hat{\alpha}'_n)'$ . The next theorem confirms that such approximation bears an  $o_P(1)$  approximation error and derives the asymptotic distribution of  $\{\hat{\theta}_n\}_{n \in \mathbb{N}}$  based on the approximation.

THEOREM 4.4: *Suppose that Assumptions 1–3 hold. Then*

$$n^{1/2}(\hat{\theta}_n - \theta_0) = -\frac{1}{R(\beta_0, 0)} K^{-1} C' n^{-1/2} \sum_{t=1}^n (1(U_t < 0) - \tau) Z_t + o_P(1),$$

and

$$D^{-1/2} n^{1/2} (\hat{\theta}_n - \theta_0) \overset{A}{\rightsquigarrow} N(0, I_1),$$

where  $D \equiv K^{-1} C' V C K^{-1}$ ,  $K = R(\beta^*, 0)^{-1} J_{\beta\gamma} J_{\gamma\gamma}^{-1} J_{\gamma\beta}$  (as introduced earlier), and  $V \equiv R(\beta_0, \alpha_0, 0)^{-2} E[Z_1 Z_1']$ .

To estimate the asymptotic covariance matrix  $D$  consistently, we need to estimate  $V$ ,  $K$ , and  $C$  consistently. For consistent estimation of  $V$ , we can use its sample analogue,  $\hat{V}_n \equiv \hat{R}_n(\hat{\beta}_n, \hat{\alpha}_n, 0)^{-2} n^{-1} \sum_{t=1}^n Z_t Z_t'$ .

On the other hand,  $K$  and  $C$  are more complicated, depending on  $J$ , the Hessian of  $R$ . The Hessian of  $\hat{R}_n$  is zero at each point in  $\mathbf{B} \times \mathbb{R}^{k_1} \times \mathbb{R}^k$ , at which it is differentiable. This rules out estimation of  $J$  by using of the Hessian of  $\hat{R}_n$ .

A way to overcome the difficulty in estimation of  $K$  and  $C$  is to employ the numerical differentiation approach described in Newey and McFadden (1994, Section 7.3). Because  $-K$  is the Hessian of  $(\beta, \alpha) \mapsto \log Q(\beta, \alpha) : \mathbf{B} \times \mathbb{R}^{k_1} \rightarrow \mathbb{R}$  at  $(\beta'_0, \alpha'_0)'$ ,  $-1$  times a second-order numerical derivative of  $\log \hat{Q}(\beta, \alpha, \cdot)$  at  $(\hat{\beta}'_n, \hat{\alpha}'_n)'$  is our estimator of  $K$ . Let  $e_i^m$  denote the unit vector along the  $i$ th axis of the Cartesian coordinate system in  $\mathbb{R}^m$ . Assume:

ASSUMPTION 4: *The sequence  $\{h_n\}_{n \in \mathbb{N}}$  consists of positive (possibly random) numbers such that  $h_n \rightarrow 0$  and  $n^{1/2}h_n \rightarrow \infty$  in probability- $P$ .*

Then our estimator of  $K$  is  $\hat{K}_n$ , whose  $(i, j)$ -element is given by

$$\begin{aligned} \hat{K}_{nij} \equiv & -\frac{1}{4h_n^2} (\log \hat{Q}_n(\hat{\theta}_n + h_n e_i + h_n e_j, \cdot) - \log \hat{Q}_n(\hat{\theta}_n - h_n e_i + h_n e_j, \cdot) \\ & - \log \hat{Q}_n(\hat{\theta}_n + h_n e_i - h_n e_j, \cdot) + \log \hat{Q}_n(\hat{\theta}_n - h_n e_i - h_n e_j, \cdot)), \quad (i, j) \in \{1, \dots, (g + k_1)\}^2, n \in \mathbb{N}. \end{aligned}$$

For  $C$ , we utilize the result of Lemma 4.2, which suggests that perturbation in  $\hat{\theta}_n = (\hat{\beta}'_n, \hat{\alpha}'_n)'$  would change  $\hat{\gamma}_n$  approximately by  $C$  times the change in  $\hat{\theta}_n$ . Let  $\hat{\gamma}_{ni}(\theta)$  denote the  $i$ th element in the usual quantile regression estimator in regression of  $y_t - (Y'_t, Z'_{t,1})\theta$  on  $Z_t$  ( $i \in \{1, 2, \dots, k\}$ ). Then our estimator of  $C$  is  $\hat{C}_n$  whose  $(i, j)$ -element is given by

$$\hat{C}_{nij} \equiv \frac{1}{2h_n} (\hat{\gamma}_{ni}(\hat{\theta}_n + h_n e_j) - \hat{\gamma}_{ni}(\hat{\theta}_n - h_n e_j)), \quad i \in \{1, \dots, k\}, j \in \{1, \dots, (g + k_1)\}.$$

Given the estimators of  $K$  and  $C$ , we estimate  $D$  by  $\hat{D}_n \equiv \hat{K}_n^+ \hat{C}'_n \hat{V}_n \hat{C}_n \hat{K}_n^+$ ,  $n \in \mathbb{N}$ , where  $\hat{K}_n^+$  is the Moore-Penrose (MP) inverse of  $\hat{K}_n$  (we use the MP inverse instead of the regular inverse to ensure that this estimator is well defined for every realization).

THEOREM 4.5: *Suppose that Assumptions 1–4 hold. Then:*

- (a)  $\{\hat{K}_n\}_{n \in \mathbb{N}}$  is weakly consistent for  $K$ .
- (b)  $\{\hat{C}_n\}_{n \in \mathbb{N}}$  is weakly consistent for  $C$ .

(c)  $\{\hat{V}_n\}_{n \in \mathbb{N}}$  is weakly consistent for  $V$ .

(d)  $\{\hat{D}_n\}_{n \in \mathbb{N}}$  is weakly consistent for  $D$ .

Remark. The same step size  $h_n$  is used in each element of  $\hat{K}_n$  and  $\hat{C}_n$  just for simplicity. One could use a different step size for each element in  $\hat{K}_n$  and  $\hat{C}_n$  without affecting the consistency results in Theorem 4.5, as long as the step size satisfies the requirements in Assumption 4.

## 5 Testing on the Regression Coefficients under Possible Weak Identification

In a way analogous to the derivation of (2), we can show that

$$\sup_{(\beta, \alpha) \in \mathbf{B} \times \mathbb{R}^{k_1}} Q(\beta, \alpha) = \sup_{\beta \in \mathbf{B}} \frac{\inf_{\gamma \in \mathbb{R}^k} R(\beta, 0, \gamma)}{\inf_{\alpha \in \mathbb{R}^{k_1}} R(\beta, \alpha, 0)} = \sup_{\beta \in \mathbf{B}} \frac{\inf_{\gamma \in \mathbb{R}^k} \mathbb{E}[\rho_\tau(U_1 - Y_1'(\beta - \beta_0) - Z_1'\gamma)]}{\inf_{\alpha \in \mathbb{R}^{k_1}} \mathbb{E}[\rho_\tau(U_1 - Y_1'(\beta - \beta_0) - Z_{1,1}'\alpha)]}.$$

If the instruments  $Z_{1,2}$  excluded from the structural equation is poorly related to  $Y_1$ , then the ratio on the right-hand can be close to one for all  $\beta \in \mathbf{B}$ , relative to the error in approximating  $Q$  by  $\hat{Q}_n$  in the  $\rho_\tau$ -IV estimation. This essentially constitutes the situation similar to that with so-called weak instruments in the standard IV regression. This raises concerns about inference on  $\beta_0$  and  $\alpha_0$ , relying on the asymptotics in Section 4.

In the usual IV regression, the AR test is known to robust to weakness of instruments. The objective function in the  $\rho_\tau$ -IV estimator is closely related to the AR statistic. This allows us to easily adapt Anderson and Rubin's (1949) approach in our setup.

Partition  $Y_t$  and  $Z_{t,1}$  as  $Y_t \equiv (Y_t^{(1)'}, Y_t^{(2)'})'$  and  $Z_{t,1} \equiv (Z_{t,1}^{(1)'}, Z_{t,1}^{(2)'})'$ , where  $Y_t^{(i)}$  and  $Z_{t,1}^{(i)}$  are  $g^{(i)} \times 1$  and  $k_1^{(i)} \times 1$  vectors, respectively,  $i = 1, 2$  (so,  $g^{(1)} + g^{(2)} = g$ ,  $k_1^{(1)} + k_1^{(2)} = k_1$ ). Partition  $\beta_0$  and  $\alpha_0$ , analogously, as  $\beta_0 = (\beta_0^{(1)'}, \beta_0^{(2)'})' \in \mathbb{R}^{g^{(1)}} \times \mathbb{R}^{g^{(2)}}$  and  $\alpha_0 = (\alpha_0^{(1)'}, \alpha_0^{(2)'})' \in \mathbb{R}^{k_1^{(1)}} \times \mathbb{R}^{k_1^{(2)}}$ . The hypothesis we consider is that  $H_0: \beta_0^{(1)} = \bar{\beta}^{(1)}$  and  $\alpha_0^{(1)} = \bar{\alpha}^{(1)}$ , where  $\bar{\beta}^{(1)}$  and  $\bar{\alpha}^{(1)}$  are known  $g^{(1)} \times 1$  and  $k_1^{(1)} \times 1$  constant vectors, respectively. Define

$$\mathbf{B}^{(2)} \equiv \{\beta^{(2)} \in \mathbb{R}^{g^{(2)}} : (\bar{\beta}^{(1)}, \beta^{(2)}) \in \mathbf{B}\}.$$

Then our test statistic is

$$\mathcal{AR}_n \equiv -2n \log \left( \sup_{(\beta^{(2)}, \alpha^{(2)}) \in \mathbf{B}^{(2)} \times \mathbb{R}^{k_1^{(2)}}} \hat{Q}_n(\bar{\beta}^{(1)}, \beta^{(2)}, \bar{\alpha}^{(1)}, \alpha^{(2)}, \cdot) \right), \quad n \in \mathbb{N}.$$

This test statistic is simply  $(-2)n$  times the logarithmic function of the maximized objective function in the  $\rho_\tau$ -IV regression of  $y_t - Y_t^{(1)'} \bar{\beta}^{(1)} - Z_{t,1}^{(1)'} \bar{\alpha}^{(1)}$  on  $Y_t^{(2)}$  and  $Z_{t,1}^{(2)}$ , taking  $Z_t$  for instruments.

Suppose that  $Y_t^{(2)}$  is well related to  $Z_{t,2}$ , while some elements of  $Y_t^{(1)}$  may be only weakly related to  $Z_{t,2}$ . Then the  $\rho_\tau$ -IV regression involved in  $\mathcal{AR}_n$  is not subject to the weak instrument problem, because only  $Y_t^{(2)}$  and  $Z_{t,1}^{(2)}$  appear as regressors in the regression. The possible weak identification problem does not affect the behavior of  $\mathcal{AR}_n$ .

The standard large sample analysis on extremum estimation shows that

$$n^{-1} \mathcal{AR}_n \rightarrow -2 \log \left( \sup_{(\beta^{(2)}, \alpha^{(2)}) \in \mathbf{B}^{(2)} \times \mathbb{R}^{k_1^{(2)}}} Q(\bar{\beta}^{(1)}, \beta^{(2)}, \bar{\alpha}^{(1)}, \alpha^{(2)}) \right) \text{ in probability-}P.$$

Under  $H_0$ , the right-hand side of this equality is zero, because

$$\begin{aligned} \sup_{(\beta^{(2)}, \alpha^{(2)}) \in \mathbf{B}^{(2)} \times \mathbb{R}^{k_1^{(2)}}} Q(\bar{\beta}^{(1)}, \beta^{(2)}, \bar{\alpha}^{(1)}, \alpha^{(2)}) &= \sup_{(\beta^{(2)}, \alpha^{(2)}) \in \mathbf{B}^{(2)} \times \mathbb{R}^{k_1^{(2)}}} Q(\beta_0^{(1)}, \beta^{(2)}, \alpha_0^{(1)}, \alpha^{(2)}) \\ &= Q(\beta_0^{(1)}, \beta_0^{(2)}, \alpha_0^{(1)}, \alpha_0^{(2)}) = 1. \end{aligned}$$

Under the alternative, on the other hand, the function  $(\beta_0^{(2)}, \alpha_0^{(2)}) \mapsto Q(\bar{\beta}^{(1)}, \beta^{(2)}, \bar{\alpha}^{(1)}, \alpha^{(2)})$  attains one nowhere on  $\mathbf{B}_2 \times \mathbb{R}^{k_2} \rightarrow \mathbb{R}$ . Thus, our test rejects  $H_0$  if  $\mathcal{AR}_n$  is far from zero.

Let  $J_{\gamma\theta_2} = (J'_{\gamma\beta_2}, J'_{\gamma\alpha_2})'$ , where  $J_{\gamma\beta_2}$  is the last  $g^{(2)}$  rows of  $J_{\gamma,\beta}$ , and  $J_{\gamma\alpha_2}$  is the last  $k_1^{(2)}$  rows of  $J_{\gamma\alpha_2}$ . Also, let

$$J_{\theta_2\theta_2} \equiv \begin{pmatrix} J_{\beta_2\beta_2} & J_{\beta_2\alpha_2} \\ J_{\alpha_2\beta_2} & J_{\alpha_2\alpha_2} \end{pmatrix}$$

where  $J_{\beta_2\beta_2}$  is the  $g^{(2)} \times g^{(2)}$  submatrix in the southwest corner of  $J_{\beta\beta}$ ;  $J_{\beta_2\alpha_2}$  is the  $g^{(2)} \times k_1^{(2)}$  submatrix in the southwest corner of  $J_{\beta\alpha}$ ;  $J_{\alpha_2\alpha_2}$  is the  $k_1^{(2)} \times k_1^{(2)}$  submatrix in the southwest corner of  $J_{\alpha\alpha}$ ; and  $J_{\alpha_2\beta_2} = J'_{\beta_2\alpha_2}$ . Define  $C^0 \equiv -J_{\gamma\gamma}^{-1} J_{\gamma\theta_2}$  and  $L \equiv R(\beta_0, \alpha_0, 0)^{-1} J_{\theta_2\theta_2}$ .

LEMMA 5.1: *Suppose that Assumptions 1–3 hold. If in addition  $H_0$  is true, then*

$$\mathcal{AR}_n \stackrel{A}{\approx} \eta' (L^{-1} - C^0 (C^{0'} L C^0)^{-1} C^{0'}) \eta,$$

$\eta$  is a  $k \times 1$  random vector distributed with  $N(0, V)$ ,

Thus,  $\{\mathcal{AR}_n\}_{n \in \mathbb{N}}$  has a nondegenerated limiting distribution, though it is not asymptotically pivotal. Among the unknown parameters in the formula for the asymptotic distribution of  $\{\mathcal{AR}_n\}_{n \in \mathbb{N}}$ ,  $C^0$  can be consistently estimated by applying  $\{\hat{C}_n\}_{n \in \mathbb{N}}$  “under the null” (Theorem 4.5). Write  $\hat{\beta}_n^0 \equiv (\bar{\beta}^{(1)'}, \hat{\beta}_n^{(2)'})'$ ,  $\hat{\alpha}_n^0 \equiv (\bar{\alpha}^{(1)'}, \hat{\alpha}_n^{(2)'})'$ , and  $\hat{\theta}_n^0 \equiv (\hat{\beta}_n^{0'}, \hat{\alpha}_n^{0'})'$ . Then our estimator of  $C$  is  $\hat{C}_n^0$  whose  $(i, j)$ -element is equal to

$$\hat{C}_{nij}^0 \equiv \frac{1}{2h_n} (\hat{\gamma}_{ni}(\hat{\theta}_n^0 + h_n e_j) - \hat{\gamma}_{ni}(\hat{\theta}_n^0 - h_n e_j)), \quad i \in \{1, \dots, k\}, j \in \{1, \dots, (g+k_1)\}.$$

Analogously,  $V$  can be estimated by  $\hat{V}_n^0 \equiv \hat{R}_n(\hat{\beta}_n^0, \hat{\alpha}_n^0, 0)^{-2} n^{-1} \sum_{t=1}^n Z_t Z_t'$ .

The matrix  $L$  is the Hessian of  $\gamma \mapsto \log R(\beta_0, \alpha_0, \gamma) : \mathbb{R}^k \rightarrow \mathbb{R}$  at the origin. We take a second-order numerical derivative of the sample counterpart of this function to estimate  $L$ . Let  $\hat{\beta}_n^{(2)}$  and  $\hat{\alpha}_n^{(2)}$  denote the estimators of the slopes for  $Y_2^{(2)}$  and  $Z_{t,1}^{(2)}$ , respectively, in  $\rho_\tau$ -IV regression of  $y_t - Y_t^{(1)'} \bar{\beta}^{(1)} - Z_{t,1}^{(1)'} \bar{\alpha}^{(1)}$  on  $Y_t^{(2)}$  and  $Z_{t,1}^{(2)}$ , taking  $Z_t$  for instruments. Our estimator  $\hat{L}_n$  of  $L$  is thus the  $k \times k$  matrix with  $(i, j)$ -element equal to

$$\begin{aligned} \hat{L}_{nij} \equiv & \frac{1}{4h_n^2} (\log \hat{R}_n(\hat{\beta}_n^0, \hat{\alpha}_n^0, \hat{\gamma}_n^0 + h_n e_i + h_n e_j, \cdot) - \log \hat{R}_n(\hat{\beta}_n^0, \hat{\alpha}_n^0, \hat{\gamma}_n - h_n e_i + h_n e_j, \cdot) \\ & - \log \hat{R}_n(\hat{\beta}_n^0, \hat{\alpha}_n^0, \hat{\gamma}_n + h_n e_i - h_n e_j, \cdot) + \log \hat{R}_n(\hat{\beta}_n^0, \hat{\alpha}_n^0, \hat{\gamma}_n - h_n e_i - h_n e_j, \cdot)), \quad i, j = 1, 2, \dots, k, \end{aligned}$$

LEMMA 5.2: *Suppose that Assumptions 1–3 and 4 hold. If in addition  $H_0$  holds,  $\{\hat{L}_n^0\}_{n \in \mathbb{N}}$  is consistent for  $L$ .*

The limiting distribution of  $\{\mathcal{AR}_n\}_{n \in \mathbb{N}}$  is that of a positive random variable whose distribution function is positively sloped at each positive point. Let  $c(p, \tilde{C}, \tilde{L}, \tilde{V})$  denote the  $(1-\alpha)$ -quantile of  $\tilde{\eta}'(\tilde{L}^+ - \tilde{C}(\tilde{C}'\tilde{L}\tilde{C}) + \tilde{C}')\tilde{\eta}$  for each  $k \times l$  matrix  $\tilde{C}$ , each  $k \times k$  symmetric matrix  $\tilde{L}$ , and each  $k \times k$  symmetric matrix  $\tilde{V}$ , where  $\tilde{\eta}$  is a  $k \times 1$  random vector distributed with  $N(0, \tilde{V})$ , and  $p \in (0, 1)$ , where  $(a, b)$  denotes the open interval between real numbers  $a$  and  $b$ . We here propose a test that  $H_0$  if and only if  $\mathcal{AR}_n$  exceeds  $c(p, \hat{C}_n, \hat{L}_n, \hat{V}_n)$ , where  $p$  is the desired size of the test. This test has the correct asymptotic size and it is consistent, as stated in the next theorem.

THEOREM 5.3: *Suppose that Assumptions 1–4 hold. Then:*

(a) If in addition  $H_0$  holds, for each  $p \in (0, 1)$ ,  $P[\mathcal{AR}_n > c(p, \hat{C}_n^0, \hat{L}_n^0, \hat{V}_n^0)] \rightarrow p$ .

(b) Suppose instead that  $H_0$  is violated, that  $(\beta^{(2)}, \alpha^{(2)}) \mapsto Q(\bar{\beta}^{(1)}, \beta^{(2)}, \bar{\beta}^{(2)}, \alpha^{(2)}) : \mathbf{B}_2 \times \mathbb{R}^{k_1^{(2)}} \rightarrow \mathbb{R}$  is uniquely maximized at a point  $(\beta^{*(2)'}, \alpha^{*(2)'}) \in \mathbf{B}_2 \times \mathbb{R}^{k_1^{(2)}}$ , and that  $R(\bar{\beta}^{(1)}, \beta^{*(2)}, \bar{\alpha}^{(1)}, \alpha^{*(2)}, \cdot) : \mathbb{R}^k \rightarrow \mathbb{R}$  is uniquely minimized at  $\gamma^*$ . Then for each  $p \in (0, 1)$ ,  $P[\mathcal{AR}_n > c(p, \hat{C}_n^0, \hat{L}_n^0, \hat{V}_n^0)] \rightarrow 1$ .

Because each quadratic form of normal random variables can be easily rewritten as a linear combination of  $\chi^2$  random variables using the eigenvalue decomposition,  $c(p, \hat{C}_n, \hat{L}_n, \hat{V}_n)$  is the  $(1 - \alpha)$ -quantile of a linear combination of  $\chi^2$  random variables. To compute  $c(p, \hat{C}_n, \hat{L}_n, \hat{V}_n)$ , we can numerically find the  $(1 - \alpha)$ -quantile of the distribution of the linear combination, evaluating the distribution function by using Farebrother's (1984) algorithm.

One potential problem in the test using  $\mathcal{AR}_n$  described above is that when  $k_2$  is large, its power may be low, just as is the case in the Anderson-Rubin test. To overcome the difficulty, we can replace  $Z_{t,2}$  with a small number of linear combinations with them, say,  $\Pi Z_t$ , where  $\Pi$  is  $m \times k$ . Write

$$V_t \equiv y_t - Y_t^{(1)'} \bar{\beta}^{(1)} - Z_{t,1}^{(1)'} \bar{\alpha}^{(1)}.$$

Then resulting test statistic is

$$\mathcal{AR}_n(\Pi) \equiv -2 \log \left( \sup_{\beta^{(2)} \in \mathbf{B}^{(2)}} \frac{\inf_{(\gamma_1, \gamma_2) \in \mathbb{R}^{k_1} \times \mathbb{R}^m} \rho_\tau(V_t - Y_t^{(2)'} \beta^{(2)} - Z_{t,1}' \gamma_1 - (\Pi Z_t)' \gamma_2)}{\inf_{\alpha^{(2)} \in \mathbb{R}^{k_1^{(2)}}} \rho_\tau(V_t - Y_t^{(2)'} \beta^{(2)} - Z_{t,1}^{(2)'} \alpha^{(2)})} \right), \quad n \in \mathbb{N}. \quad (6)$$

In the ratio on the right-hand side of (6), the numerator is equal to

$$\inf_{(\gamma_1^{(1)}, \gamma_1^{(2)}, \gamma_2) \in \mathbb{R}^{k_1^{(1)}} \times \mathbb{R}^{k_1^{(2)}} \times \mathbb{R}^m} \rho_\tau \left( U_t - Y_t^{(1)'} (\bar{\beta}^{(1)} - \beta_0^{(1)}) - Y_t^{(2)'} (\beta^{(2)} - \beta_0^{(2)}) - Z_{t,1}^{(1)'} \gamma_1^{(1)} - Z_{t,1}^{(2)'} \gamma_1^{(2)} - (\Pi Z_t)' \gamma_2 \right)$$

while the denominator is equal to

$$\inf_{\gamma_1^{(2)} \in \mathbb{R}^{k_1^{(2)}}} \rho_\tau \left( U_t - Y_t^{(1)'} (\bar{\beta}^{(1)} - \beta_0^{(1)}) - Y_t^{(2)'} (\beta^{(2)} - \beta_0^{(2)}) - Z_{t,1}^{(1)'} (\bar{\alpha}^{(1)} - \alpha_0^{(1)}) - Z_{t,1}^{(2)'} \gamma_1^{(2)} \right).$$

This gives us intuition that when the numerator is minimized, the violation of  $\alpha_0^{(1)} = \bar{\alpha}^{(1)}$  would be captured by the nonzero value of  $\gamma_1^{(1)}$ , which should drive  $\mathcal{AR}_n(\Pi)$  away from zero, while the violation of  $\beta_0^{(1)} = \bar{\beta}^{(1)}$  would be caught by nonzero value of  $\gamma_2$ , which again should result in a  $\mathcal{AR}_n(\Pi)$  away from zero, if  $\Pi Z_t$  and  $Y_t^{(1)}$  are well related.

A practical question is how to choose  $\Pi$ . An intuitively appealing way is to estimate it by regressing each element of  $Y_t^{(1)}$  on  $Z_t$ . Nevertheless, if the estimated  $\Pi$  and the other parts of  $\mathcal{AR}_n(\Pi)$  are dependent under the null, it is difficult to set the critical value in our test.

In certain cases, we can estimate  $\Pi$  in such a way that the estimated  $\Pi$  is asymptotically independent of  $\mathcal{AR}_n(\Pi)$ , so that we can make the asymptotic size correct, treating the estimated  $\Pi$  can be treated if it were a constant. We describe such a setup in the next assumption.

ASSUMPTION 5: (a)  $g^{(1)} = g$ .

(b) *There exists a  $g \times k$  constant matrix  $\Pi_0$  such that  $Z_1$  is independent from  $U_t$  and  $\epsilon_1 \equiv Y_1 - \Pi_0 Z_1$ , and  $E[\epsilon_1] = 0$ .*

(c)  $E[|Y_1|^4 + |Z_1|^4] < \infty$ .

Note that  $\mathcal{AR}_{\Pi,n}$  is asymptotically equivalent to a quadratic form of  $n^{-1/2} \sum_{t=1}^n (1(U_t < 0) - \tau) Z_t$ . Consider the OLS regression of each element of  $Y_t^{(1)}$  on  $Z_t$  and  $(1(\hat{U}_t < 0) - \tau) Z_t$ , where  $\hat{U}_t$  is the fitted residual from the  $\rho_\tau$ -IV estimator. Then we can show that  $\hat{\Pi}_n$ , the slope estimator for  $Z_t$  is  $\sqrt{n}$ -consistent for  $\Pi_0$ , and  $n^{1/2}(\hat{\Pi}_n - \Pi_0)$  and  $n^{-1/2} \sum_{t=1}^n (1(U_t < 0) - \tau) Z_t$  are asymptotically independent. Because of this asymptotic independence, the AR test using  $(Z_{t,1}, \hat{\Pi}_n Z_t)$  for instruments treating as if  $\hat{\Pi}_n$  were a constant matrix has asymptotically the correct size.

Let  $\hat{C}_n^0(\Pi)$ ,  $\hat{L}_n^0(\Pi)$ , and  $\hat{V}_n^0(\Pi)$  denote the estimators obtained by replacing  $Z_{t,2}$  with  $\Pi' Z_t$  in the estimators  $\hat{C}_n^0$ ,  $\hat{L}_n^0$ , and  $\hat{V}_n^0$ , respectively. Also, let  $R_\Pi : \mathbf{B} \times \mathbb{R}^{k_1} \times \mathbb{R}^k \rightarrow \mathbb{R}$  and  $Q_\Pi : \mathbf{B} \times \mathbb{R}^{k_1} \rightarrow \mathbb{R}$  be the functions obtained by replacing  $Z_{1,2}$  with  $\Pi' Z_1$  in  $R$  and  $Q$ , respectively.

THEOREM 5.4: *Suppose that Assumptions 1–5 hold.*

(a) *If in addition  $H_0$  holds, for each  $p \in (0, 1)$ ,  $P[\mathcal{AR}_n(\hat{\Pi}_n) > c(p, \hat{C}_n^0(\hat{\Pi}_n), \hat{L}_n^0(\hat{\Pi}_n), \hat{V}_n^0(\hat{\Pi}_n))] \rightarrow p$ .*

(b) *Suppose instead that  $H_0$  is violated, that  $(\beta^{(2)}, \alpha^{(2)}) \mapsto Q_{\Pi_0}(\bar{\beta}^{(1)}, \beta^{(2)}, \bar{\beta}^{(2)}, \alpha^{(2)}) : \mathbf{B}_2 \times \mathbb{R}^{k_1^{(2)}} \rightarrow \mathbb{R}$  is uniquely maximized at a point  $(\beta^{*(2)'}, \alpha^{*(2)'}) \in \mathbf{B}_2 \times \mathbb{R}^{k_1^{(2)}}$ , and that  $R_{\Pi_0}(\bar{\beta}^{(1)}, \beta^{*(2)}, \bar{\alpha}^{(1)}, \alpha^{*(2)}, \cdot) : \mathbb{R}^k \rightarrow \mathbb{R}$  is uniquely minimized at  $\gamma^*$ . Then for each  $p \in (0, 1)$ ,  $P[\mathcal{AR}_n > c(p, \hat{C}_n^0, \hat{L}_n^0, \hat{V}_n^0)] \rightarrow 1$ .*

## 6 Monte Carlo simulations

[Coming soon]

## 7 Concluding Remarks

[Coming soon]

## Appendix A Mathematical Proofs

Given Assumption 1, write  $\|\xi\|_{\rho_\tau} \equiv \rho_\tau(\xi)$  for each  $\xi \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ . Then  $\|\cdot\|_{\rho_\tau}$  is a pseudo norm on  $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ . Using  $\|\cdot\|_{\rho_\tau}$ ,  $R$  can be written as

$$R(\beta, \alpha, \gamma) = \|y_1 - Y_1' \beta - Z_{1,1}' \alpha - Z_1' \gamma\|_{\rho_\tau}, \quad (\beta, \alpha, \gamma) \in \mathbf{B} \times \mathbb{R}^{k_1} \times \mathbb{R}^k.$$

It follows that the minimization in the numerator of the ratio on the right-hand side of (3) is the  $\|\cdot\|_{\rho_\tau}$ -metric projection of  $y_1 - Y_1' \beta$  on  $Z_1$ , while the minimization in the denominator is the  $\|\cdot\|_{\rho_\tau}$ -metric projection of  $y_1 - Y_1' \beta$  on  $Z_{1,1}$ . The norm  $\|\cdot\|_{\rho_\tau}$  is closely related to the  $\mathcal{L}_1$  norm  $\|\cdot\|_1$ . They actually generate the equivalent topologies, because

$$\|\xi\|_{\rho_\tau} \leq \|\xi\|_1 \leq \frac{1}{\min\{\tau, 1 - \tau\}} \|\xi\|_{\rho_\tau}$$

One (but not only) important implication of the equivalence is that  $\|\xi\|_{\rho_\tau} = 0$  if and only if  $\|\xi\|_1 = 0$ . Our analysis uses the equivalence of the two norms, mostly without mentioning it explicitly.

We show below that  $\{\sup_{\alpha \in \mathbb{R}^{k_1}} \hat{Q}_n(\beta, \alpha, \cdot)\}_{n \in \mathbb{N}}$  converges to  $\sup_{\alpha \in \mathbb{R}^{k_1}} Q(\beta, \alpha)$  uniformly in  $\beta$  on the compact set  $\mathbf{B}$ . We can then conclude that  $\{\hat{\beta}_n\}_{n \in \mathbb{N}}$  is consistent for  $\beta_0$ , because  $\beta_0$  is the unique maximizer of  $\sup_{\alpha \in \mathbb{R}^{k_1}} Q(\beta, \alpha)$  on  $\mathbf{B}$ . Once the consistency  $\hat{\beta}_n$  is established, we can also prove that  $\{\hat{\alpha}_n\}_{n \in \mathbb{N}}$  at which  $R(\beta_0, \cdot) : \mathbb{R}^{k_1} \rightarrow \mathbb{R}$  is minimized, converges a.s.- $P$  to  $\alpha_0$ , by utilizing the convexity of  $\hat{R}_n(\hat{\beta}_n, \alpha, \cdot)$  in  $\alpha$  and the pointwise convergence of  $\{\hat{R}_n(\hat{\beta}_n, \alpha, \cdot)\}_{n \in \mathbb{N}}$  to  $R(\beta_0, \alpha)$  for each  $\alpha$ .

We first establish a few lemmas. For later conveniences, some lemmas have more generality than we need for proving Theorem 4.1. The generality will be useful in our proof of 4.4.

LEMMA A.1: *Suppose that Assumptions 1 holds. Then for each  $\beta \in \mathbf{B}$ ,*

$$\inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(\beta, 0, \gamma, \cdot) - \inf_{\gamma \in \mathbb{R}^k} R(\beta, 0, \gamma) \rightarrow 0 \text{ a.s.-}P,$$

and

$$\inf_{\alpha \in \mathbb{R}^{k_1}} \hat{R}_n(\beta, \alpha, 0, \cdot) - \inf_{\alpha \in \mathbb{R}^{k_1}} R(\beta, \alpha, 0) \rightarrow 0 \text{ a.s.-}P.$$

Proof of LEMMA A.1: The two convergence results can be proved in similar manners. We only prove the first one. Let  $\beta$  be an arbitrary point in  $\mathbf{B}$ . Then the  $\|\cdot\|_{\rho_\tau}$ -metric projection of  $y_1 - Y_1'\beta$  on the linear subspace spanned by  $Z_1$  exists and is in general a compact set. By the linear independence of  $Z_1$  (Assumption 1), this further means that  $\Gamma_1 \equiv \arg \min_{\gamma \in \mathbb{R}^k} R(\beta, 0, \gamma)$  is compact. It follows that there exists a closed ball  $\Gamma_2$  containing  $\Gamma_1$  in its interior. Now fix a point  $\gamma_1$  in  $\Gamma_1$ . By the Kolmogorov law of large numbers (Rao 1973, p. 115),  $\{\hat{R}_n(\beta, 0, \gamma_1, \cdot)\}_{n \in \mathbb{N}}$  converges to  $R(\beta, 0, \gamma_1)$  a.s.- $P$ . Also, by Jennrich's uniform law of large numbers (Jennrich 1969, Theorem 2),  $\{\hat{R}_n(\beta, 0, \gamma, \cdot) - R(\beta, 0, \gamma)\}_{n \in \mathbb{N}}$  converges to zero uniformly in  $\gamma$  on the boundary  $\partial\Gamma_2$  of  $\Gamma_2$  a.s.- $P$ . Because  $\hat{R}_n(\beta, 0, \gamma, \cdot)$  is convex as a function of  $\gamma$ , and

$$R(\beta, 0, \gamma_1) < \inf_{\gamma \in \partial\Gamma_2} R(\beta, 0, \gamma),$$

it follows from the above-mentioned facts that

$$\hat{R}_n(\beta, 0, \gamma_1, \cdot) < \inf_{\gamma \in \mathbb{R}^k \setminus \Gamma_2} \hat{R}_n(\beta, 0, \gamma, \cdot)$$

for almost all  $n \in \mathbb{N}$  a.s.- $P$ . On the other hand, by Jennrich's uniform law of large numbers,  $\{\hat{R}_n(\beta, 0, \gamma, \cdot) - R(\beta, 0, \gamma)\}_{n \in \mathbb{N}}$  converges to zero uniformly in  $\gamma \in \Gamma_2$  a.s.- $P$ , so that

$$\inf_{\gamma \in \Gamma_2} \hat{R}_n(\beta, 0, \gamma, \cdot) \rightarrow R(\beta, 0, \gamma_1) \text{ a.s.-}P.$$

The desired result therefore follows.  $\square$

LEMMA A.2: *Suppose that Assumptions 1 holds. Then*

$$\sup_{\beta \in \mathbf{B}} \left| \inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(\beta, 0, \gamma, \cdot) - \inf_{\gamma \in \mathbb{R}^k} R(\beta, 0, \gamma) \right| \rightarrow 0 \text{ a.s.-}P,$$

and

$$\sup_{\beta \in \mathbf{B}} \left| \inf_{\alpha \in \mathbb{R}^{k_1}} \hat{R}_n(\beta, \alpha, 0, \cdot) - \inf_{\alpha \in \mathbb{R}^{k_1}} R(\beta, \alpha, 0) \right| \rightarrow 0 \text{ a.s.-}P.$$

Proof of LEMMA A.2: We only prove the first convergence result, as the second one can be shown in an analogous manner. Because Lemma A.1 has shown the corresponding pointwise a.s. convergence, and  $\mathbf{B}$  is compact, it suffices to show that the series in question is strongly stochastically equicontinuous (Andrews 1992, Theorem 2).

Let  $\beta_1$  and  $\beta_2$  be arbitrary points in  $\mathbf{B}$ . Also, let  $g_{nj}$  be Koenker and Bassett's (1978) estimator in  $\tau$ -quantile regression of  $y_t - Y_t'\beta_j$  on  $Z_t$ , i.e.,

$$\hat{R}_n(\beta_j, 0, g_{nj}, \cdot) = \inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(\beta_j, 0, \gamma, \cdot),$$

for  $j = 1, 2$ . Then we have that for each  $n \in \mathbb{N}$ ,

$$\begin{aligned} & \hat{R}_n(\beta_1, 0, g_{n1}, \cdot) - \hat{R}_n(\beta_2, 0, g_{n2}, \cdot) \\ &= (\hat{R}_n(\beta_1, 0, g_{n1}, \cdot) - \hat{R}_n(\beta_1, 0, g_{n2}, \cdot)) + (\hat{R}_n(\beta_1, 0, g_{n2}, \cdot) - \hat{R}_n(\beta_2, 0, g_{n2}, \cdot)) \\ &\leq \hat{R}_n(\beta_1, 0, g_{n2}, \cdot) - \hat{R}_n(\beta_2, 0, g_{n2}, \cdot), \end{aligned}$$

where the inequality holds, because  $\hat{R}_n(\beta_1, 0, g_{n1}, \cdot) \leq \hat{R}_n(\beta_1, 0, g_{n2}, \cdot)$  for each  $n \in \mathbb{N}$ . We further have that

$$\begin{aligned} \hat{R}_n(\beta_1, 0, g_{n2}, \cdot) - \hat{R}_n(\beta_2, 0, g_{n2}, \cdot) &= n^{-1} \sum_{t=1}^n (\rho_\tau(y_t - Y_t'\beta_1 - Z_t'g_{n2}) - \rho_\tau(y_t - Y_t'\beta_2 - Z_t'g_{n2})) \\ &\leq n^{-1} \sum_{t=1}^n |Y_t'\beta_1 - Y_t'\beta_2| \leq |\beta_1 - \beta_2| n^{-1} \sum_{t=1}^n |Y_t|. \end{aligned}$$

It follows that for each  $n \in \mathbb{N}$

$$\hat{R}_n(\beta_1, 0, g_{n1}, \cdot) - \hat{R}_n(\beta_2, 0, g_{n2}, \cdot) \leq |\beta_1 - \beta_2| n^{-1} \sum_{t=1}^n |Y_t|.$$

Analogously, we can also show that for each  $n \in \mathbb{N}$

$$\hat{R}_n(\beta_2, 0, g_{n2}, \cdot) - \hat{R}_n(\beta_1, 0, g_{n1}, \cdot) \leq |\beta_1 - \beta_2| n^{-1} \sum_{t=1}^n |Y_t|.$$

Thus, it holds that for each  $n \in \mathbb{N}$

$$\left| \inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(\beta_2, 0, \gamma, \cdot) - \inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(\beta_1, 0, \gamma, \cdot) \right| = |\hat{R}_n(\beta_2, 0, g_{n2}, \cdot) - \hat{R}_n(\beta_1, 0, g_{n1}, \cdot)| \leq |\beta_1 - \beta_2| n^{-1} \sum_{t=1}^n |Y_t|.$$

Because  $\{n^{-1} \sum_{t=1}^n |Y_t|\}_{n \in \mathbb{N}}$  converges to  $E[|Y_1|]$  a.s.- $P$  by the Kolmogorov strong law of large numbers, the desired result follows by Andrews (1992, Lemma 2).  $\square$

LEMMA A.3: *Suppose that Assumptions 1 and 2 hold. For each  $\beta \in \mathbf{B}$ ,  $\{\inf_{\alpha \in \mathbb{R}^k} \hat{Q}_n(\beta, \alpha, \cdot)\}$  converges to  $\inf_{\alpha \in \mathbb{R}^k} Q(\beta, \alpha)$  uniformly in  $\beta \in \mathbf{B}$  a.s.- $P$*

Proof of LEMMA A.3: Because the linear independence of the elements of  $X_1 = (y_1, Y_1', Z_1')$  in Assumption 1 implies that for each  $\beta \in \mathbf{B}$ , the distance between  $y_1 - Y_1'\beta$  and the  $\|\cdot\|_{\rho_\tau}$ -metric projection of  $y_1 - Y_1'\beta$  on the subspace spanned by  $Z_1$  is positive, i.e.,  $\inf_{\alpha \in \mathbb{R}^{k_1}} R(\beta, \alpha, 0) > 0$ . Because  $\beta \mapsto \inf_{\alpha \in \mathbb{R}^{k_1}} R(\beta, \alpha, 0) : \mathbf{B} \rightarrow \mathbb{R}$  is continuous, it is bounded away from zero on  $\mathbf{B}$ . The desired results from this fact and Lemma A.2, because  $(r_1, r_2) \mapsto r_1/r_2 : \mathbb{R} \times (a, \infty) \rightarrow \mathbb{R}$  with  $a > 0$  is a Lipschitz function.  $\square$

LEMMA A.4: *Suppose that Assumptions 1 and 2 hold. Let  $\{b_n\}_{n \in \mathbb{N}}$  be a sequence of  $\mathbf{B}$ -valued random vectors on  $(\Omega, \mathcal{F}, P)$  converging to  $\beta_0$  a.s.- $P$  (in probability- $P$ ). Let  $\{a_n\}_{n \in \mathbb{N}}$  be sequences of  $k_1 \times 1$  vectors on  $(\Omega, \mathcal{F}, P)$  satisfying that for each  $n \in \mathbb{N}$ ,  $\hat{R}_n(b_n, a_n, 0, \cdot) = \inf_{\alpha \in \mathbb{R}^{k_1}} \hat{R}_n(b_n, \alpha, 0, \cdot)$ . Then:*

(a) *Then  $a_n \rightarrow \alpha_0$  a.s.- $P$  (in probability- $P$ ).*

(b) *Let  $\{c_n\}_{n \in \mathbb{N}}$  be a sequence of  $k \times 1$  random vectors on  $(\Omega, \mathcal{F}, P)$  satisfying that for each  $n \in \mathbb{N}$   $\hat{R}_n(b_n, a_n, c_n, \cdot) = \inf_{\gamma \in \mathbb{R}^k} \hat{R}_n(b_n, a_n, \gamma, \cdot)$ . Then  $c_n \rightarrow 0$  a.s.- $P$  (in probability- $P$ ), provided that the minimizer of  $R(\beta_0, \alpha_0, \cdot) : \mathbb{R}^k \rightarrow \mathbb{R}$  over  $\mathbb{R}^k$  is unique.*

Proof of LEMMA A.4: We only prove the result for  $\{a_n\}_{n \in \mathbb{N}}$ . The result for  $\{c_n\}_{n \in \mathbb{N}}$  can be established in an analogous way.

Suppose that  $b_n \rightarrow \beta_0$  a.s.- $P$ . Then for each  $\alpha \in \mathbb{R}^{k_1}$ ,  $\{\hat{R}_n(b_n, \alpha, 0, \cdot)\}_{n \in \mathbb{N}}$  converges to  $R(\beta_0, \alpha, 0)$  a.s.- $P$ , because for each  $\alpha \in \mathbb{R}^{k_1}$ ,  $\{\hat{R}_n(\beta, \alpha, 0, \cdot)\}_{n \in \mathbb{N}}$  converges to  $R(\beta, \alpha, 0)$  uniformly in  $\beta \in \mathbf{B}$  by Jennrich's uniform law of large numbers (Jennrich 1969, Theorem 2). Further, we can apply Rockafellar (1970, Theorem 10.8) to show that the convergence is uniform in  $\alpha$  over any compact subset of  $\mathbb{R}^{k_1}$ , because for each  $n \in \mathbb{N}$ ,  $\hat{R}_n(b_n, \alpha, 0, \cdot)$  is convex in  $\alpha$  over  $\mathbb{R}^{k_1}$ .

Take an arbitrary compact subset  $\mathbf{A}_1$  of  $\mathbb{R}^{k_1}$  that contain  $\alpha_0$  in its interior. Then  $\{\hat{R}_n(b_n, \alpha_0, 0, \cdot)\}$  converges to  $R(\beta_0, \alpha_0, 0)$  a.s.- $P$ ;  $\{\hat{R}_n(b_n, \alpha, 0, \cdot)\}$  converges to  $R(\beta_0, \alpha, 0)$  uniformly on  $\alpha \in \partial \mathbf{A}_1$  a.s.- $P$ ; and  $R(\beta_0, \alpha_0, 0) < \inf_{\alpha \in \partial \mathbf{A}_1} R(\beta_0, \alpha, 0)$ , because  $\alpha_0$  is the unique minimizer of  $R(\beta_0, \cdot, 0)$  on  $\mathbb{R}^{k_1}$  by Assumption 2.

Because  $\hat{R}_n(b_n, \alpha, 0, \cdot)$  is convex in  $\alpha$ , it follows that

$$\hat{R}_n(b_n, \alpha_0, 0, \cdot) < \inf_{\alpha \in \mathbb{R}^{k_1} \setminus \mathbf{A}_1} \hat{R}_n(b_n, \alpha, 0, \cdot)$$

for almost all  $n \in \mathbb{N}$  a.s.- $P$ . That is,  $a_n \in A_1$  for almost all  $n \in \mathbb{N}$  a.s.- $P$ . Because  $A_1$  is an arbitrary compact subset containing  $\alpha_0$  in its interior, this establishes the a.s.- $P$  convergence of  $\{a_n\}_{n \in \mathbb{N}}$  to  $\alpha_0$ . The convergence of  $\{a_n\}_{n \in \mathbb{N}}$  in probability in the current lemma immediately follows from the result of the a.s. convergence of  $\{a_n\}_{n \in \mathbb{N}}$  by using the subsequence theorem.  $\square$

Proof of THEOREM 4.1: By Assumption 2,  $\beta \mapsto \sup_{\alpha \in \mathbb{R}^k} Q(\beta, \alpha) : \mathbf{B} \rightarrow \mathbb{R}$  is uniquely maximized at  $\beta_0$ . Because  $\hat{\beta}_n$  maximizes  $\sup_{\alpha \in \mathbb{R}^k} \hat{Q}_n(\beta, \alpha, \cdot)$  with respect to  $\beta$  over the compact subset  $\mathbf{B}$ , and  $\{\sup_{\alpha \in \mathbb{R}^k} \hat{Q}_n(\beta, \alpha, \cdot)\}_{n \in \mathbb{N}}$  converges to  $\sup_{\alpha \in \mathbb{R}^k} Q(\beta, \alpha)$  uniformly in  $\beta \in \mathbf{B}$  a.s.- $P$ , it follows by Pötscher and Prucha (1991, Lemma 4.2) that  $\{\hat{\beta}_n\}_{n \in \mathbb{N}}$  converges to  $\beta_0$  a.s.- $P$ . Further, applying Lemma A.4(a) by setting  $b_n = \hat{\beta}_n$  and  $a_n = \hat{\alpha}_n$  establishes that the strong consistency of  $\hat{\alpha}_n$  for  $\alpha_0$ . The result therefore follows.  $\square$

In proving Lemmas 4.2, 4.3 and Theorem 4.4, we use the following lemma.

LEMMA A.5: *Suppose that Assumptions 1–3 hold, and let  $\{\tilde{d}_{nj} \equiv (b'_{nj}, a_{nj}, g'_{nj})' : \Omega \rightarrow \mathbf{B} \times \mathbb{R}^{k_1} \times \mathbb{R}^k\}_{n \in \mathbb{N}}$  be a sequence of random vectors that converges to  $d_0 \equiv (\beta'_0, \alpha'_0, 0_{1 \times k})'$ ,  $j = 1, 2$ . Then*

$$\begin{aligned} & \hat{R}_n(b_{n2}, a_{n2}, g_{n2}, \cdot) - \hat{R}_n(b_{n1}, a_{n1}, g_{n1}, \cdot) \\ &= -n^{-1} \sum_{t=1}^n (\tau - 1(U_t < 0)) \tilde{X}'_t(\tilde{d}_{n2} - \tilde{d}_{n1}) + \frac{1}{2}(\tilde{d}_{n2} - d_0)' J(\tilde{d}_{n2} - d_0) - \frac{1}{2}(\tilde{d}_{n1} - d_0)' J(\tilde{d}_{n1} - d_0) \\ & \quad + o_P(n^{-1/2}(|d_{n2} - d_{n1}| + |d_{n1} - d_0|^2 + |d_{n2} - d_0|^2)), \end{aligned} \tag{7}$$

where  $\tilde{X}_t \equiv (Y_t, Z'_{t,1}, Z'_t)'$ ,  $t \in \mathbb{N}$ .

Proof of LEMMA A.5: Define  $r : \mathbb{R} \times \mathbb{R}^{g+k_1+k} \times \mathbb{R}^{g+k_1+k} \times \mathbb{R}^{g+k_1+k} \rightarrow \mathbb{R}$  by

$$\begin{aligned} r(y, \tilde{x}, d_1, d_2) &\equiv \frac{1}{|d_2 - d_1|} \left( \rho_\tau(y - \tilde{x}'d_2) - \rho_\tau(y - \tilde{x}'d_1) + (\tau - 1(y - \tilde{x}'d_0 < 0)) \tilde{x}'(d_2 - d_1) \right), \\ & (y, \tilde{x}, d_1, d_2) \in \mathbb{R} \times \mathbb{R}^{g+k_1+k} \times \mathbb{R}^{g+k_1+k} \times \mathbb{R}^{g+k_1+k}, \end{aligned}$$

with the rule that division by zero is zero. Also, following Pollard (1985), let  $\nu_n$  denote the standardized sample average operator such that for each function  $f : \mathbb{R} \times \mathbb{R}^{l+k} \rightarrow \mathbb{R}$  with  $\mathbb{E}[|f(Y_1, \tilde{X}_1)|] < \infty$

$$\nu_n f(\cdot, \cdot) = n^{-1/2} \sum_{t=1}^n (f(Y_t, \tilde{X}_t) - \mathbb{E}[f(Y_1, \tilde{X}_1)]), \quad n \in \mathbb{N}.$$

By the definition of  $r$ , we obtain that

$$\begin{aligned} \hat{R}_n(d_2, \cdot) - \hat{R}_n(d_1, \cdot) &= R(d_2) - R(d_1) - \left( \ell + n^{-1} \sum_{t=1}^n (\tau - 1(U_t < 0)) \tilde{X}_t \right)' (d_2 - d_1) \\ &\quad + n^{-1/2} |d_2 - d_1| \nu_n r(\cdot, \cdot, d_1, d_2) \end{aligned}$$

for each  $(d_1, d_2) \in \mathbb{R}^{l+k} \times \mathbb{R}^{l+k}$ , where  $\ell$  is the gradient of  $R$  at  $(\beta_0, \alpha_0, 0_{1 \times k})$ , which is equal to  $-\mathbb{E}[(\tau - 1(U_1 < 0)) \tilde{X}_1]$ . Taking the second-order Taylor expansion of  $R(d_1)$  and  $R(d_2)$  about  $d_0$  on the right-hand side of this equality and replacing  $d_1$  with  $\tilde{d}_{n1}$  and  $d_2$  with  $\tilde{d}_{n2}$  in the resulting equality yields the desired result, if  $\{\nu_n r(\cdot, \cdot, \tilde{d}_{n1}, \tilde{d}_{n2})\}_{n \in \mathbb{N}}$  converges to zero in probability- $P$ . It thus suffices to show the convergence of  $\{\nu_n r(\cdot, \cdot, \tilde{d}_{n1}, \tilde{d}_{n2})\}$  to zero in probability- $P$ .

It is straightforward to verify that  $r(y_1, \tilde{X}_1, \theta_1, \theta_2) \leq 2|\tilde{X}_1|$ , from which it follows that

$$\mathbb{E} \left[ \sup_{(d_1, d_2) \in \mathbb{R}^{g+k_1+k} \times \mathbb{R}^{g+k_1+k}} r(y_1, \tilde{X}_1, d_1, d_2)^2 \right] \leq 4\mathbb{E}[|\tilde{X}_1|^2] < \infty.$$

Also,  $\{r(\cdot, \cdot, d_1, d_2) : (d_1, d_2) \in \mathbb{R}^{g+k_1+k} \times \mathbb{R}^{g+k_1+k}\}$  can be expressed as a sum of a fixed member of functions from a polynomial class. These facts imply that  $\{\nu_n r(\cdot, \cdot, d_1, d_2)\}_{n \in \mathbb{N}}$  is stochastically equicontinuous at  $(d_0, d_0)$  (Pollard 1985, pp. 311–312). Further,  $r(y_1, \tilde{X}_1, d_1, d_2)^2$  converges to zero as  $(d_1, d_2) \rightarrow (d_0, d_0)$  a.s.- $P$ , and  $r(y_1, \tilde{X}_1, d_1, d_2)^2$  is dominated by  $4|\tilde{X}_1|^2$  with a finite moment. It follows by the dominated convergence theorem that  $\mathbb{E}[r(y_1, \tilde{X}_1, d_1, d_2)^2] \rightarrow 0$  as  $(d_1, d_2) \rightarrow (d_0, d_0)$ .

Now let  $\{\mathbf{U}_n \subset \mathbb{R}^{g+k_1+k} \times \mathbb{R}^{g+k_1+k}\}_{n \in \mathbb{N}}$  be an arbitrary sequence of balls centered at  $(d'_0, d'_0)'$  that shrinks down to  $(d'_0, d'_0)'$ . Then, as Pollard (1985, page. 309) explains, it follows from the above-mentioned facts that  $\sup_{(d_1, d_2) \in \mathbf{U}_n} |\nu_n r(\cdot, \cdot, d_1, d_2)| \rightarrow 0$  in probability- $P$ . Thus,  $\{\nu_n r(\cdot, \cdot, \tilde{d}_{n1}, \tilde{d}_{n2})\}$  converges to zero in probability- $P$ , given that  $\{\tilde{d}_{nj}\}_{n \in \mathbb{N}}$  converges to  $d_0$  in probability- $P$ ,  $j = 1, 2$ .  $\square$

LEMMA A.6: *Let  $(\Omega, \mathcal{F}, P)$  be a probability space. Suppose that a sequence of random vectors  $\{\eta_n : \Omega \rightarrow \mathbb{R}^m\}_{n \in \mathbb{N}}$  and a sequence of random variables  $\{\xi_n : \Omega \rightarrow \mathbb{R}\}_{n \in \mathbb{N}}$  satisfy that  $-\eta_n' A \eta_n + \xi_n \geq 0$  for each  $n \in \mathbb{N}$ , where  $A$  is a positive definite  $m \times m$  symmetric matrix. Also, let  $\{\zeta_n : \Omega \rightarrow \mathbb{R}\}_{n \in \mathbb{N}}$  be a sequence of random variables. Suppose that  $\xi_n = o_P(|\eta_n| + |\eta_n|^2 + |\zeta_n|)$  as  $n \rightarrow \infty$ . Then  $|\eta_n| = o_P(|\zeta_n|^{1/2} + 1)$  as  $n \rightarrow \infty$ .*

We now prove Lemma 4.2.

Proof of LEMMA 4.2: The existence of  $\{c_n\}$  follows immediately from the fact that the minimization of

$\hat{R}_n(b_n, a_n, \gamma, \cdot)$  in terms of  $\gamma$  is the  $\|\cdot\|_{\rho_\tau}$ -metric projection of  $(y_1 - Y_1' b_n - Z_{1,1}' a_n, y_2 - Y_2' b_n - Z_{1,2}' a_n, \dots, y_n - Y_n' b_n - Z_{1,n}' a_n)$  on the space spanned by the rows of  $(Z_1, Z_2, \dots, Z_n)$ .

To prove the second result, we first show that  $\{c_n\}$  converges to 0 in probability- $P$ , and then apply Lemmas A.5 and A.6. For each fixed  $\gamma \in \mathbb{R}^k$ ,  $\hat{R}_n(\beta, \alpha, \gamma, \cdot)$  is convex in  $\beta$  and  $\alpha$ . By the Kolmogorov strong law of large numbers and Hjort and Pollard (1993, Lemma 1),  $\{\hat{R}_n(\beta, \alpha, \gamma, \cdot)\}_{n \in \mathbb{N}}$  converges to  $R(\beta, \alpha, \gamma)$  uniformly in  $(\beta', \alpha')$  in each neighborhood of  $(\beta_0', \alpha_0')$  in probability- $P$ . Because  $\{(b_n', a_n')'\}_{n \in \mathbb{N}}$  converges to  $(\beta_0', \alpha_0')$  in probability- $P$  by hypothesis, it follows that  $\{\hat{R}_n(b_n, a_n, \gamma, \cdot)\}_{n \in \mathbb{N}}$  converges to  $R(\beta_0, \alpha_0, \gamma)$  for each  $\gamma \in \mathbb{R}^k$ . Under Assumptions 1-3(c), this fact implies by Hjort and Pollard (1993, Lemma 2) that  $\{c_n\}$  converges to 0 in probability- $P$ .

We now set  $b_n$  to both  $b_{n1}$  and  $b_{n2}$ ,  $c_n$  to  $g_{n1}$ ,

$$g_n^\dagger \equiv C \begin{pmatrix} b_n - \beta_0 \\ a_n - \alpha_0 \end{pmatrix} + J_{\gamma\gamma}^{-1} n^{-1} \sum_{t=1}^n (\tau - 1(U_t < 0)) Z_t.$$

to  $g_{n2}$  in (7) and multiply the resulting equality by  $n$  to obtain that

$$\begin{aligned} 0 &\leq n(\hat{R}_n(b_n, a_n, g_n^\dagger) - \hat{R}_n(b_n, a_n, c_n)) \\ &= -\frac{1}{2} n^{1/2} (c_n - g_n^\dagger)' J_{\gamma\gamma} n^{1/2} (c_n - g_n^\dagger) \\ &\quad + o_P(n^{1/2} |c_n - g_n^\dagger| + n|b_n - \beta_0|^2 + n|a_n - \alpha_0|^2 + n|c_n|^2 + n|g_n^\dagger|^2) \\ &= -\frac{1}{2} n^{1/2} (c_n - g_n^\dagger)' J_{\gamma\gamma} n^{1/2} (c_n - g_n^\dagger) \\ &\quad + o_P(n^{1/2} |c_n - g_n^\dagger|) + n|c_n - g_n^\dagger|^2 + n|b_n - \beta_0|^2 + n|a_n - \alpha_0|^2 + 1, \end{aligned}$$

where the second equality holds because  $|g_n^\dagger| = O_P(|b_n - \beta_0| + |a_n - \alpha_0| + 1)$  and  $c_n = O_P(|c_n - g_n^\dagger| + |g_n^\dagger|)$ .

The result follows from this inequality by Lemma A.6.  $\square$

Proof of LEMMA 4.3: Let  $\{c_n\}_{n \in \mathbb{N}}$  be as in Lemma 4.2. Note that  $\{\hat{R}_n(b_n, a_n, c_n)\}_{n \in \mathbb{N}}$  and  $\{\hat{R}_n(b_n, 0)\}_{n \in \mathbb{N}}$

converge to zero in probability- $P$ . Applying the delta method with this fact, we obtain that

$$\begin{aligned} n \log \hat{Q}_n(b_n, a_n, \cdot) &= n(\log \hat{R}_n(b_n, a_n, c_n) - \log \hat{R}_n(b_n, a_n, 0)) \\ &= \frac{1}{R(\beta_0, \alpha_0, 0)} n(\hat{R}_n(b_n, a_n, c_n) - \hat{R}_n(b_n, a_n, 0)) \end{aligned} \quad (8)$$

$$\begin{aligned} &- \frac{1}{2R(\beta_0, \alpha_0, 0)^2} n(\hat{R}_n(b_n, a_n, c_n) - R(\beta_0, \alpha_0, 0))^2 \\ &+ \frac{1}{2R(\beta_0, \alpha_0, 0)^2} n(\hat{R}_n(b_n, a_n, 0) - R(\beta_0, \alpha_0, 0))^2 \\ &+ o_P(n(\hat{R}_n(b_n, a_n, c_n) - R(\beta_0, \alpha_0, 0))^2 + n(\hat{R}_n(b_n, a_n, 0) - R(\beta_0, \alpha_0, 0))^2). \end{aligned} \quad (9)$$

We apply Lemma A.5 to each of the non-remainder terms on the right-hand side of this equality:

$$\begin{aligned} &n(\hat{R}_n(b_n, a_n, c_n, \cdot) - \hat{R}_n(b_n, a_n, 0)) \\ &= -\frac{1}{2}n^{1/2}c'_n J_{\gamma\gamma} n^{1/2}c_n + o_P(n^{1/2}|c_n| + n|b_n - \beta_0|^2 + n|a_n - \alpha_0|^2 + |c_n|^2) \\ &= -\frac{1}{2}n^{1/2}c'_n J_{\gamma\gamma} n^{1/2}c_n + o_P(n^{1/2}|b_n - \beta_0| + n^{1/2}|a_n - \alpha_0| + n|b_n - \beta_0|^2 + n|a_n - \alpha_0|^2 + 1), \end{aligned}$$

$$\begin{aligned} &n^{1/2}(\hat{R}_n(b_n, a_n, c_n) - R(\beta_0, \alpha_0, 0)) \\ &= n^{1/2}(\hat{R}_n(b_n, a_n, c_n) - \hat{R}_n(\beta_0, \alpha_0, 0)) + n^{1/2}(\hat{R}_n(\beta_0, \alpha_0, 0) - R(\beta_0, \alpha_0, 0)) \\ &= -(n^{-1} \sum_{t=1}^n (\tau - 1(U_t < 0)) (Y'_t, Z'_{t,1})')' n^{1/2}(\tilde{\theta}_n - \theta_0) + n^{1/2}(\hat{R}_n(\beta_0, \alpha_0, 0) - R(\beta_0, \alpha_0, 0)) \\ &\quad + o_P(n^{1/2}|b_n - \beta_0| + n^{1/2}|a_n - \alpha_0| + n|b_n - \beta_0|^2 + n|a_n - \alpha_0|^2 + 1), \end{aligned}$$

and

$$\begin{aligned} &n^{1/2}(\hat{R}_n(b_n, a_n, 0) - R(\beta_0, \alpha_0, 0)) \\ &= n^{1/2}(\hat{R}_n(b_n, a_n, 0) - \hat{R}_n(\beta_0, \alpha_0, 0)) + n^{1/2}(\hat{R}_n(\beta_0, \alpha_0, 0) - R(\beta_0, \alpha_0, 0)) \\ &= -\left(n^{-1} \sum_{t=1}^n (\tau - 1(U_t < 0)) (Y'_t, Z'_{t,1})'\right)' n^{1/2}(b_n - \beta_0) + n^{1/2}(\hat{R}_n(\beta_0, \alpha_0, 0) - R(\beta_0, \alpha_0, 0)) \\ &\quad + o_P(n^{1/2}|b_n - \beta_0| + n^{1/2}|a_n - \alpha_0| + n|b_n - \beta_0|^2 + n|a_n - \alpha_0|^2 + 1). \end{aligned}$$

Substituting these into (8) and applying Lemma 4.2 yields the desired result.  $\square$

Proof of THEOREM 4.4: Let

$$\tilde{\theta}_n^\dagger \equiv \theta_0 - \frac{1}{R(\beta_0, 0)} K^{-1} C' n^{-1} n^{-1} \sum_{t=1}^n (\tau - 1(U_t < 0)) Z_t, \quad n \in \mathbb{N},$$

and let  $b_n^\dagger$  and  $a_n^\dagger$  denote the vectors containing the first  $g$  elements and the remaining elements of  $\tilde{\theta}_n$ , respectively. Then, by Lemma 4.3, we have that

$$\begin{aligned} 0 &\leq n \log \hat{Q}_n(\hat{\beta}_n, \hat{\alpha}_n, \cdot) - n \log \hat{Q}_n(b_n^\dagger, a_n^\dagger, \cdot) \\ &= \frac{1}{2} n^{1/2} (\hat{\theta}_n - \tilde{\theta}_n^\dagger)' K n^{1/2} (\hat{\theta}_n - \tilde{\theta}_n^\dagger) \\ &\quad + o_P(n^{1/2} |\hat{\beta}_n - \beta_0| + n^{1/2} |\hat{\alpha}_n - \alpha_0| + n |\hat{\beta}_n - \beta_0|^2 + n |\hat{\alpha}_n - \alpha_0|^2 + 1). \end{aligned}$$

The first result follows from this equality by Lemma A.6. For the second result, apply the central limit theorem (CLT) for i.i.d. random vectors (Rao 1973, p. 128) to show that  $\{n^{-1/2} \sum_{t=1}^n (\tau - 1(U_t < 0)) Z_t\}_{n \in \mathbb{N}}$  is asymptotically distributed with  $N(0, V)$ , and then apply the continuous mapping theorem.  $\square$

Proof of THEOREM 4.5: To prove (a), let  $\{\tilde{\theta}_n^\dagger\}_{n \in \mathbb{N}}$  be as in the proof of Theorem 4.4 and  $\{\delta_n\}_{n \in \mathbb{N}}$  an arbitrary sequence of  $(g + k_1) \times 1$  random vectors that converges to the origin in probability- $P$ . Recall that the first three terms on the right-hand side of (5) is minimized when  $(b_n', a_n')' \tilde{\theta}_n^\dagger$  and that  $\{n^{1/2}(\hat{\theta}_n - \tilde{\theta}_n^\dagger)\}_{n \in \mathbb{N}}$  converges to zero in probability- $P$  by Theorem 4.4. Using these facts with Lemma 4.3, we can show that  $n \log \hat{Q}_n(\hat{\theta}_n, \cdot) - n \log \hat{Q}_n(\tilde{\theta}_n^\dagger, \cdot) = o_P(1)$  and

$$\begin{aligned} n \log \hat{Q}_n(\hat{\theta}_n + \delta_n, \cdot) - n \log \hat{Q}_n(\tilde{\theta}_n^\dagger, \cdot) &= -\frac{1}{2} n^{1/2} (\hat{\theta}_n - \tilde{\theta}_n^\dagger + \delta_n)' K n^{1/2} (\hat{\theta}_n - \tilde{\theta}_n^\dagger + \delta_n) + o_P(n^{1/2} |\delta_n| + |\delta_n|^2 + 1) \\ &= -\frac{1}{2} n^{1/2} \delta_n' K n^{1/2} \delta_n + o_P(n^{1/2} |\delta_n| + |\delta_n|^2 + 1) \end{aligned} \tag{10}$$

By taking each of  $\tau_n e_i + \tau_n e_j$ ,  $-\tau_n e_i + \tau_n e_j$ ,  $\tau_n e_i - \tau_n e_j$ , and  $-\tau_n e_i - \tau_n e_j$  for  $\delta_n$  in this equality and using the resulting equality in the definition of  $\hat{K}_{nij}$  ( $i, j = 1, 2, \dots, l$ ), we obtain that

$$4n\tau_n^2 \hat{K}_{nij} = 4\tau_n^2 K_{ij} + o_P(n^{-1/2} \tau_n + \tau_n^2 + 1).$$

Dividing both sides of this equality by  $4n\tau_n^2$  and applying Assumption 4 yields the desired result.

To prove (b), let  $\delta$  be an arbitrary  $(g + k_1) \times 1$  vector. By Lemma 4.2, we have that for each  $i = 1, 2, \dots, k$  and each  $j = 1, 2, \dots, l$ ,  $\hat{\gamma}_{ni}(\hat{\theta}_n + \tau_n \delta) - \hat{\gamma}_{ni}(\hat{\theta}_n - \tau_n \delta) = 2\tau_n C \delta_n + o_P(\tau_n)$ . It follows that

$$\frac{1}{2\tau_n} (\hat{\gamma}_n(\hat{\beta}_n + \tau_n \delta) - \hat{\gamma}_n(\hat{\beta}_n - \tau_n \delta)) = C \delta + o_P(1).$$

Taking  $e_j$  for  $\delta$  for each  $j = 1, 2, \dots, k$  in this equality completes the proof.

To show (c), notice that  $\{\hat{R}_n(\beta, \alpha, 0)\}_{n \in \mathbb{N}}$  converges to  $R(\beta, \alpha, 0)$  uniformly in  $(\beta', \alpha)'$  in each closed ball centered at  $(\beta'_0, \alpha'_0)'$  in probability- $P$ , because  $\{\hat{R}_n(\beta, \alpha, 0)\}$  converges to  $R(\beta, \alpha, 0)$  in probability- $P$  for each  $(\beta, \alpha) \in \mathbf{B} \times \mathbb{R}^{k_1}$ , and  $\hat{R}_n(\beta, \alpha, 0)$  is convex in  $\beta$  and  $\alpha$  for each  $n \in \mathbb{N}$  (Hjort and Pollard 1993, Lemma 1). It follows by Theorem 4.4 that  $\{\hat{R}_n(\hat{\beta}_n, \hat{\alpha}_n, 0, \cdot)\}_{n \in \mathbb{N}}$  converges to  $R(\beta_0, \alpha_0, 0)$ . Also,  $n^{-1} \sum_{t=1}^n Z_t Z_t'$  converges to  $E[Z_1 Z_1']$  a.s.- $P$  by the Kolmogorov strong law of large numbers and Hjort and Pollard (1993, Lemma 1). The desired result follows from these by the Slutsky theorem. Finally, the claim (d) immediately follows from a-c by the Slutsky theorem.  $\square$

Proof of LEMMA 5.1: Using Lemma 4.3 and Theorem 4.4 the  $\rho_\tau$ -IV regression of  $y_t - Y_t^{(1)'} \bar{\beta}^{(1)} - Z_{t,1}^{(1)'} \bar{\alpha}^{(1)}$  on  $Y_t^{(2)}$  and  $Z_{t,1}^{(2)}$ , taking  $Z_t$  for instruments, we obtain that

$$\begin{aligned} \mathcal{AR}_n &= \frac{1}{R(\beta_0, \alpha_0, 0)} \left( n^{-1/2} \sum_{t=1}^n (\tau - 1(U_t < 0)) Z_t \right)' (L^{-1} - C^0 K^{0-1} C^{0'}) \\ &\quad \times \frac{1}{R(\beta_0, \alpha_0, 0)} \left( n^{-1/2} \sum_{t=1}^n (\tau - 1(U_t < 0)) Z_t \right) + o_P(1), \end{aligned}$$

where  $K^0 \equiv R(\beta_0, \alpha_0, 0)^{-1} J_{\theta_2 \gamma} J_{\gamma \gamma}^{-1} J_{\gamma \theta_2}$ . Because  $R(\beta_0, \alpha_0, 0)^{-1} n^{-1/2} \sum_{t=1}^n (\tau - 1(U_t < 0)) Z_t \stackrel{A}{\sim} \eta$  by the central limit theorem (CLT) for i.i.d. random vectors (Rao 1973, p. 128), it follows by the continuous mapping theorem and the asymptotic equivalence lemma that

$$\mathcal{AR}_n \stackrel{A}{\sim} \eta' (L^{-1} - C^0 K^{0-1} C^{0'}) \eta.$$

Applying the fact that  $K^0 = C^{0'} L C^0$  to this result completes the proof.  $\square$

Proof of LEMMA 5.2: Let  $\{\delta_n\}_{n \in \mathbb{N}}$  be an arbitrary sequence of  $k \times 1$  random vectors that converges to the origin in probability- $P$ . By Lemma A.5 with Theorem 4.4 and Lemma 4.2, we have that

$$\hat{R}_n(\hat{\beta}_n^0, \hat{\alpha}_n^0, \hat{\gamma}_n^0 + \delta_n, \cdot) - \hat{R}_n(\hat{\beta}_n^0, \hat{\alpha}_n^0, \hat{\gamma}_n^0, \cdot) = \frac{1}{2} \delta_n' J_{\gamma \gamma} \delta_n + o_P \left( n^{-1/2} |\delta_n| + |\delta_n|^2 + n^{-1} \right).$$

By the delta method, it follows that

$$\begin{aligned} &\log \hat{R}_n(\hat{\beta}_n^0, \hat{\alpha}_n^0, \hat{\gamma}_n^0 + \delta_n, \cdot) - \log \hat{R}_n(\hat{\beta}_n^0, \hat{\alpha}_n^0, \hat{\gamma}_n^0, \cdot) \\ &= \frac{1}{R(\beta_0, \alpha_0, 0)} \left( \hat{R}_n(\hat{\theta}_n, \hat{\gamma}_n^0 + \delta_n, \cdot) - \hat{R}_n(\hat{\theta}_n, \hat{\gamma}_n^0, \cdot) \right) = \frac{1}{2} \delta_n' L \delta_n + o_P(1). \end{aligned}$$

The desired result follows from this equality in essentially the same way as Theorem 4.5(a) is proved using (10).  $\square$

Proof of THEOREM 5.3: To prove (a), it suffices to show that for each  $p \in (0, 1)$ ,  $\{c(p, \hat{C}_n, \hat{L}_n, \hat{V}_n)\}_{n \in \mathbb{N}}$  is weakly consistent for  $c(p, C, L, V)$ . For each  $k \times l$  matrix  $\tilde{C}$ , each  $k \times k$  symmetric matrix  $\tilde{L}$ , and each  $k \times k$  symmetric matrix  $\tilde{V}$ , let  $F(\cdot, \tilde{C}, \tilde{L}, \tilde{V}) : \mathbb{R} \rightarrow \mathbb{R}$  denote the distribution function of  $\tilde{\eta}'(\tilde{L}^+ - \tilde{C}(\tilde{C}'\tilde{L}\tilde{C})^+ \tilde{C}')\tilde{\eta}$  with  $\tilde{\eta}$  distributed with  $N(0, \tilde{V})$ . Also, define  $\{F_n : \mathbb{R} \times \Omega \rightarrow \mathbb{R}\}_{n \in \mathbb{N}}$  and  $F^* : \mathbb{R} \rightarrow \mathbb{R}$  by  $\hat{F}_n(x, \cdot) \equiv F(x, \hat{C}_n, \hat{V}_n, \hat{L}_n)$ ,  $x \in \mathbb{R}$ ,  $n \in \mathbb{N}$ , and  $F^*(x) \equiv F(x, C, L, V)$ ,  $x \in \mathbb{R}$ . Because  $F^*$  is positively sloped at  $c(p, C, L, V)$  for each  $p \in (0, 1)$ , the convergence of  $\{c(p, \hat{C}_n, \hat{L}_n, \hat{V}_n)\}_{n \in \mathbb{N}}$  in probability- $P$  to  $c(p, C, L, V)$  for each  $p \in (0, 1)$  is implied by the convergence of  $\{\hat{F}_n(x, \cdot)\}_{n \in \mathbb{N}}$  in probability- $P$  to  $F^*(x)$  for each  $x \in \mathbb{R}$ . We below prove this sufficient condition.

Notice that the characteristic function  $\phi(\cdot, \tilde{C}, \tilde{L}, \tilde{V})$  of  $F(\cdot, \tilde{C}, \tilde{L}, \tilde{V})$  is given by

$$\phi(s, \tilde{C}, \tilde{L}, \tilde{V}) \equiv \det(I - 2is\tilde{V}^{1/2}(\tilde{L}^+ - \tilde{C}(\tilde{C}'\tilde{L}\tilde{C})^+ \tilde{C}')\tilde{V}^{1/2}),$$

where  $i$  denotes the imaginary unit. Because  $\hat{C}_n \rightarrow C$ ,  $\hat{L}_n \rightarrow L$ , and  $\hat{V}_n \rightarrow V$  in probability- $P$ , it follows from the subsequent theorem (Lukacs 1975, Theorem 2.4.4) that for each subsequence  $\{n_i\}_{i \in \mathbb{N}}$  of  $\{n\}_{n \in \mathbb{N}}$ , there exists a further subsequence  $\{n_{i_j}\}_{j \in \mathbb{N}}$  of it, along which  $\hat{C}_{n_{i_j}} \rightarrow C$ ,  $\hat{L}_{n_{i_j}} \rightarrow L$ , and  $\hat{V}_{n_{i_j}} \rightarrow V$  as  $j \rightarrow \infty$  a.s.- $P$ . Because  $V$  is positive definite, we have that

$$\phi(s, \hat{C}_{n_{i_j}}, \hat{L}_{n_{i_j}}, \hat{V}_{n_{i_j}}) = \det \hat{V}_{n_{i_j}} \det(\hat{V}_{n_{i_j}}^+ - 2is(\hat{L}_{n_{i_j}}^+ - \hat{C}_{n_{i_j}}(\hat{C}'_{n_{i_j}}\hat{L}_{n_{i_j}}\hat{C}_{n_{i_j}})^+ \hat{C}'_{n_{i_j}})), \quad s \in \mathbb{R}$$

for each  $s \in \mathbb{R}$  and a.a.  $n \in \mathbb{N}$  a.s.- $P$ . By the continuity theorem (Billingsley 1968, Theorem 7.6), this implies that for each  $x \in \mathbb{R}$ ,  $\{F_{n_{i_j}}(x, \cdot)\}_{j \in \mathbb{N}}$  converges to  $F^*(x)$  a.s.- $P$ . By the subsequence theorem, it follows from this fact that for each  $x \in \mathbb{R}$ ,  $\{F_n(x, \cdot)\}_{n \in \mathbb{N}}$  converges to  $F^*(x)$  in probability- $P$ . The desired result therefore follows.

For (b), it suffices to show that  $c(p, \hat{C}_n, \hat{L}_n, \hat{V}_n) = O_P(1)$ , because for each  $\bar{c} \in \mathbb{R}$ ,  $P[\mathcal{AR}_n > \bar{c}] \rightarrow 1$  if  $(\beta^{(2)}, \alpha^{(2)}) \mapsto Q(\bar{\beta}^{(1)}, \beta^{(2)}, \bar{\alpha}^{(1)}, \alpha^{(2)}) : \mathbf{B}_2 \times \mathbb{R}^{k_1^{(2)}} \rightarrow \mathbb{R}$  attains one nowhere on  $\mathbf{B}_2 \times \mathbb{R}^{k_1^{(2)}}$ .

We only give a sketch of the proof here. It is possible to derive results similar to those of Lemmas 4.2, 4.3, and Theorem 4.4 without imposing  $H_0$  in the  $\rho_\tau$ -IV regression of  $y_t - Y_t^{(1)'}\bar{\beta}^{(1)} - Z_{t,1}^{(1)'}\bar{\alpha}^{(1)}$  on  $Y_t^{(2)}$  and  $Z_{t,1}^{(2)}$ , taking  $Z_t$  for instruments, by employing essentially the same techniques. Although the constants in the resulting formulas are different from those found in Lemmas 4.2, 4.3, and Theorem 4.4, essentially the same logic delivers the convergence of  $\{\hat{C}_n^0\}_{n \in \mathbb{N}}$ ,  $\{\hat{L}_n^0\}_{n \in \mathbb{N}}$ , and  $\{\hat{V}_n^0\}_{n \in \mathbb{N}}$  in probability- $P$ . The stochastic limit of  $\{\hat{L}_n^0\}_{n \in \mathbb{N}}$  is the Hessian of  $R(\bar{\beta}^{(1)}, \beta^{*(2)}, \bar{\alpha}^{(1)}, \alpha^{*(2)}, \cdot) : \mathbb{R}^k \rightarrow \mathbb{R}$  at  $\gamma^*$ , which is nonsingular by

Assumption 3(c). Using the same argument as in the proof of (a), we can show that  $\{c(p, \hat{C}_n, \hat{L}_n, \hat{V}_n)\}_{n \in \mathbb{N}}$  converges in probability- $P$ , so that it is  $O_P(1)$ .  $\square$

## REFERENCES

- Anderson, T. W. and Rubin, H., 1949. Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations, *Annals of Mathematical Statistics*, 20, 46–63.
- Andrews, D. W. K., 1992. Generic Uniform Convergence, *Econometric Theory*, 8, 241–257.
- Andrews, D. W. K., Moreira, M. J., and Stock, J. H., 2004. Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression, *Econometrica*, 74, 715–752.
- Billingsley, P., 1968. *Convergence of Probability Measures*, Wiley Series in Probability and Mathematical Statistics. Wiley, New York, NY.
- Bound, J., Jaeger, D. A., and Baker, R. M., 1995. Problems with Instrumental Variables Estimation When the Correlation Between The Instruments and the Endogenous Explanatory Variable is Weak, *Journal of the American Statistical Association*, 90, 443–450.
- Chernozhukov, V. and Hansen, C., 2001. An IV Model of Quantile Treatment Effects, Working Paper 02-06, Massachusetts Institute of Technology, Department of Economics, Cambridge, MA.
- Chernozhukov, V. and Hansen, C., 2006. Instrumental Quantile Regression Inference for Structural and Treatment Effect Models, *Journal of Econometrics*, 132, 491–525.
- Chernozhukov, V. and Hansen, C., 2008. Instrumental Variable Quantile Regression: A Robust Inference Approach, *Journal of Econometrics*, 142, 379–398.
- Choi, I. and Phillips, P. C. B., 1992. Asymptotic and Finite Sample Distribution Theory for IV Estimator and Tests in Partially Identified Structural Equations, *Journal of Econometrics*, 51, 113–150.
- Dufour, J.-M., 1997. Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models, *Econometrica*, 65, 1365–1387.

- Farebrother, R. W., 1984. The Distribution of a Positive Linear Combination of Chi-Square Random Variables, *Applied Statistics – Journal of the Royal Statistical Society, Series C*, 33, 332–339.
- Gallant, A. R. and White, H., 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, New York.
- Hillier, G., 1990. On the Normalization of Structural Equations: Properties of Direction Estimators, *Econometrica*, 58, 1181–1194.
- Hjort, N. L. and Pollard, D., 1993. *Asymptotics for Minimizers of Convex Processes*, Preprint, Department of Statistics, Yale University.
- Jennrich, R. I., 1969. Asymptotic Properties of Non-Linear Least Squares Estimators, *Annals of Mathematical Statistics*, 40, 633–643.
- Jun, S. J., 2008. Weak Identification Robust Tests in an Instrumental Quantile Model, *Journal of Econometrics*, 144, 118–138.
- Kleibergen, F., 2002. Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression, *Econometrica*, 70, 1781–1803.
- Kleibergen, F., 2005. Testing Parameters in GMM without Assuming That They Are Identified, *Econometrica*, 73, 1103–1123.
- Koenker, R. W. and Bassett, Jr., G. W., 1978. Regression Quantiles, *Econometrica*, 46, 33–50.
- Lukacs, E., 1975. *Stochastic Convergence*, vol. 30 of Probability and Mathematical Statistics Series. Academic Press, New York.
- Moreira, M. J., 2001. Tests With Correct Size When Instruments Can Be Arbitrarily Weak, Working Paper 37, Center for Labor Economics, University of California, Berkeley.
- Moreira, M. J., 2003. A Conditional Likelihood Ratio Test for Structural Models, *Econometrica*, 71, 1027–1048.
- Nelson, C. R. and Startz, R., 1990a. The Distribution of the Instrumental Variables Estimator and Its  $t$ -Ratio When the Instrument is a Poor One, *Journal of Business*, 63, S125–S140.

- Nelson, C. R. and Startz, R., 1990b. Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator, *Econometrica*, 58, 967–976.
- Newey, W. K. and McFadden, D., 1994. Large Sample Estimation and Hypothesis Testing, in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 36, pp. 2111–2245. North-Holland, New York.
- Phillips, P. C. B., 1984. The Exact Distribution of LIML: I, *International Economic Review*, 25, 249–261.
- Phillips, P. C. B., 1985. The Exact Distribution of LIML: II, *International Economic Review*, 26, 21–36.
- Pollard, D., 1985. New Ways to Prove Central Limit Theorems, *Econometric Theory*, 1, 295–313.
- Pötscher, B. M. and Prucha, I. R., 1991. Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and Approximation Concepts, *Econometric Reviews*, 10, 125–216.
- Rao, C. R., 1973. *Linear Statistical Inference and Its Applications*, Wiley Series in Probability and Mathematical Statistics. Wiley, New York, NY, 2nd edn.
- Rockafellar, R. T., 1970. *Convex Analysis*, vol. 28 of Princeton Mathematical Series. Princeton University Press, Princeton, NJ.
- Sakata, S., 1997. Instrumental Variable Estimation Based on Mean Absolute Deviation, mimeo, University of Michigan.
- Sakata, S., 2007. Instrumental Variable Estimator Based on Conditional Median Restriction, *Journal of Econometrics*, 141, 350–382.
- Sargan, J. D., 1983. Identification and Lack of Identification, *Econometrica*, 51, 1605–1633.
- Staiger, D. and Stock, J. H., 1997. Instrumental Variables Regression with Weak Instruments, *Econometrica*, 65, 557–586.
- Stock, J. H. and Wright, J. H., 2000. GMM with Weak Identification, *Econometrica*, 68, 1055–1096.