BOUNDING THE DIFFERENCE BETWEEN TRUE AND NOMINAL REJECTION PROBABILITIES IN TESTS OF HYPOTHESES ABOUT INSTRUMENTAL VARIABLES MODELS

by

Joel L. Horowitz
Department of Economics
Northwestern University
Evanston, IL 60208

June 2020

ABSTRACT

This paper presents a simple method for carrying out inference in a wide variety of possibly nonlinear IV models under weak assumptions. The method provides a finite-sample bound on the difference between the true and nominal probabilities of rejecting a correct null hypothesis. The method is a non-Studentized version of the S test of Stock and Wright (2000) but is implemented and analyzed differently. It does not require restrictive distributional assumptions, linearity of the estimated model, simultaneous equations, or information about whether the instruments are strong or weak. It can be applied to quantile IV models that may be nonlinear and can be used to test a parametric IV model against a nonparametric alternative. It provides information about the relation between the "degree of weakness" of the instruments and the power of the test. The bound presented here holds in finite samples, regardless of the strength of the instruments.

Key Words: Instrumental variables, normal approximation, finite-sample bounds

JEL Listing: C21, C26

I thank Eric Auerbach, Ivan Canay, Xiaohong Chen, Xu Cheng, Denis Chetverikov, Whitney Newey, and Vladimir Spokoiny for helpful comments and discussions, and Caleb Kwon for research assistance.

BOUNDING THE DIFFERENCE BETWEEN TRUE AND NOMINAL REJECTION PROBABILITIES IN TESTS OF HYPOTHESES ABOUT INSTRUMENTAL VARIABLES MODELS

1. INTRODUCTION

This paper presents a simple, easily implemented method for carrying out inference in a wide variety of possibly nonlinear IV models under weak assumptions. It is applicable to linear, nonlinear, and quantile IV models and does not require knowledge of whether the instruments are strong or weak. The method provides a finite-sample bound on the difference between the true and nominal probabilities of rejecting a correct null hypothesis. The method is a non-Studentized version of the S test of Stock and Wright (2000) but is implemented and analyzed differently. It can also be interpreted as a method for testing the hypothesis that a multivariate mean is zero.

A hypothesis H_0 about a finite-dimensional parameter θ can be tested by using a test statistic that is a quadratic form in the sample analog of the identifying moment conditions. Examples of such statistics include the Anderson-Rubin (1949) statistic; the S statistic of Stock and Wright (2000); and the subset Anderson-Rubin statistic of Guggenberger, Kleibergen, Mavroeidis, and Chen (2012); and Guggenberger, Kleibergen, and Mavroeidis (2019). Except in special cases, the finite-sample distributions of these statistics and the statistic presented here are complicated functions of the unknown population distribution of the observed variables. We overcome this problem for our statistic by approximating the unknown population distribution with a normal distribution. The finite-sample distribution of the resulting approximate test statistic can be computed by simulation with any desired accuracy. We obtain a finite-sample bound on the difference between the true and nominal probabilities of rejecting a correct H_0 (the error in the rejection probability or ERP) when the critical value is obtained by using the approximate test statistic. We also obtain a finite-sample lower bound on the probability of rejecting a false H_0 (power of the test). We do not obtain the exact finite sample distribution of the test statistic. Methods for exact finite-sample inference in IV estimation depend on strong assumptions about the population from which the data are sampled or the model being estimated. The method described in this paper does not make such assumptions.

Advantages of the method include:

1. It is easy to understand and implement. Simplicity is one of the method's main advantages.

2. It applies to a wide variety of linear and nonlinear models, including mean and quantile IV models.

3. It permits subvector inference.

4. It can be used to test a parametric mean or quantile IV model against a nonparametric alternative.

5. It does not require identification of the parameter about which inference is made.

6. It does not require strong distributional assumptions, simultaneous equations, knowledge of whether the instruments are weak, or testing for weakness.

7. It provides a finite-sample upper bound on the difference between the true and nominal probabilities of rejecting a correct null hypothesis (error in the rejection probability or ERP).

8. It provides a finite-sample lower bound on the power of the test.

9. It provides information about the relation between the "degree of weakness" of the instruments and the power of the test.

Many other methods have some of these features, but we are not aware of another method that has all of them, including simplicity. The method described here does not provide a test whose exact finite-sample size and power are known or that has the optimality properties of some other tests. It is not possible for a test to have a known finite-sample size and power and also to apply to a class of models and distributions as broad as the class considered here. It is possible, however, to obtain a finite-sample upper bound on the ERP and a finite-sample lower bound on the power.

The normal approximation used here is a multivariate generalization of the Berry-Esséen theorem due to Bentkus (2003) and modifed by Raič (2019). Other normal approximations have been developed by Chernozhukov, Chetverikov, and Kato (2017) and Spokoiny and Zhilova (2015), among many others. Chernozhukov, Chetverikov, and Kato (2013) and Spokoiny and Zhilova (2015) provide reviews. The error of Bentkus's (2003) approximation converges to zero more rapidly as the sample size increases than the errors of the other approximations when the number of instruments and exogenous covariates is small compared to the sample size.

Section 2 of this paper reviews the related literature. Section 3 describes the version of the standard IV model that we consider, the hypotheses that are tested, and the test method. Section 4 presents the main result for the model of Section 3. Section 5 presents extensions to quantile IV models and to testing a parametric model against a nonparametric alternative. Section 6 presents the results of a Monte Carlo investigation of the numerical performance of the method. Section 7 presents two empirical applications, and Section 8 presents conclusions. The proofs of theorems are presented in the appendix, which is Section 9.

2. LITERATURE REVIEW

Horowitz (2006) and Horowitz and Lee (2009) describe asymptotic tests of mean and quantile IV models against nonparametric alternatives. Santos (2012) describes a test for a partially identified setting.

These authors obtain the asymptotic distributions of their test statistics. The results presented in this paper include finite-sample bounds on the differences between the true and nominal probabilities of rejecting correct and incorrect null hypotheses.

Jun (2008) and Andrews and Mikusheva (2016) describe asymptotic tests for quantile IV models that are robust to weak instruments. Other asymptotic tests of for quantile IV models can be based on any estimation method that yields an estimator of the model's parameters that is asymptotically normally distributed after suitable centering and scaling. Chernozhukov, Hansen, and Jansson (2009) describe an exact finite-sample test of a hypothesis about a parameter in a class of parametric quantile IV models. The test presented in this paper applies to both mean and quantile IV models and can also be used to test a parametric mean or quantile IV model against a nonparametric alternative.

Inference based on IV estimates is especially problematic if the instruments are weak. With weak instruments, conventional asymptotic approximations can be highly inaccurate. Nelson and Startz (1990a, 1990b) illustrate this problem with a simple model. See, also, Angrist and Krueger (1991); Bound, Jaeger, and Baker (1995); and Hansen, Hausman, and Newey (2008). An important feature of the method presented in this paper is that it does not require special treatment of weak instruments. There is a long history of research aimed at developing reliable methods for inference in IV estimation in conditional mean models, and the associated literature is very large. One stream of research has been concerned with deriving the exact finite-sample distributions of IV estimators and test statistics based on IV estimators. The test of Anderson and Rubin (1949) is a well-known early example of this research. Phillips (1983) and the references therein present additional results of early research in this stream. Recent examples of exact finite-sample results include Andrews and Marmer (2008); Andrews, Moreira, and Stock (2006); Dufour and Taamouti (2005); Moreira (2003, 2009); and Andrews, Marmer, and Yu (2019).

Another stream of research derives non-standard or higher order asymptotic approximations to the distributions of IV estimators and test statistics. Holly and Phillips (1979); Rothenberg (1984); Staiger and Stock (1997), Kitamura and Stutzer (1997); Imbens, Spady, and Johnson (1998); Wang and Zivot (1998); Stock and Wright (2000), Newey and Smith (2004); Guggenberger and Smith (2005); Andrews and Cheng (2012); Carrasco and Tchuente (2016); Cheng (2015); Andrews and Mikusheva (2016), among others, are examples of research in this stream.

A third stream of research aims at deriving the asymptotic distributions of estimators and test statistics when the number of instruments is an increasing function of the sample size and, with most methods, the instruments may be weak. Andrews and Stock (2007a) review much of this literature. Examples include Bekker (1994); Kleibergen (2002); Andrews, Moreira, and Stock (2006); Andrews and

Soares (2007); Andrews and Stock (2007b); Hansen, Hausman, and Newey (2008); Newey and Windmeijer (2009); and Andrews and Guggenberger (2018).

3. THE STANDARD IV MODEL, HYPOTHESES, AND METHOD

3.1 *The Model and Hypotheses*

The model considered in this this section and Section 4 is

(3.1) $Y = g(X,\theta) + U; \quad E(U \mid Z) = 0$,

where Y is a scalar outcome variable, X is a vector of covariates, U is a scalar random variable, g is a known real-valued function, and θ is an unknown finite-dimensional vector of constant parameters. The parameter θ is contained in a parameter set $\Theta \subset \mathbb{R}^d$ for some $d \geq 1$. One or more components of X may be endogenous. Z is a vector of instruments for X. The elements of Z include any exogenous components of X. U can have any (possibly unknown) form of heteroskedasticity that is consistent with (3.1) and the regularity conditions given in Section 4. Let q denote the dimension of Z. The dimension of X does not enter the notation used in this paper.

Let $\{Y_i, X_i, Z_i : i = 1,...,n\}$ be an independent random sample from the distribution of (Y, X, Z). Let Z_{ij} ($i = 1,...,n;\ j = 1,...,q$) denote the j'th component of Z_i. For any $\theta \in \Theta$, define

$$T_n(\theta) = n^{-1} \sum_{j=1}^{q} \left\{ \sum_{i=1}^{n} Z_{ij} [Y_i - g(X_i, \theta)] \right\}^2 .$$

Denote the covariance matrix of the random vector $Z[Y - g(X,\theta)]$ by $\Sigma(\theta)$.

We consider two hypotheses about θ, one simple and one composite. The simple null hypothesis is

(3.2) $H_0 : \theta = \theta_0$

for some $\theta_0 \in \Theta$ against the alternative

$H_1 : \theta \neq \theta_0$.

Under hypothesis (3.2), $\Sigma(\theta_0) = E(ZZ'U^2)$. The matrix $E(ZZ'U^2)$ will be denoted by Σ without the argument θ_0 when this will not cause confusion.

To describe the composite null hypothesis, let ϑ be a subvector of θ, and let $\theta = (\vartheta', \beta')'$. The composite null hypothesis is

(3.3) $H_0 : \vartheta = \vartheta_0$.

The alternative hypothesis is

$H_1 : \vartheta \neq \vartheta_0$.

For the composite hypothesis, define $\mathcal{B} = \{b : (\vartheta_0', b')' \in \Theta\}$ and $\hat{\theta} = \arg\min\{T_n(\theta) : \theta \in \mathcal{B}\}$. A hypothesis about a linear combination of components of θ can be put into the form (3.2) or (3.3) by redefining the components of θ and, therefore, does not require a separate formulation.

3.2 *Test Statistics*

The statistic for testing the simple null hypothesis (3.2) is $T_n(\theta_0)$. Let $c_\alpha(\theta_0)$ denote the α-level critical value for testing the simple hypothesis $H_0: \theta = \theta_0$. That is, $c_\alpha(\theta_0)$ is the $1-\alpha$ quantile of the distribution of $T_n(\theta_0)$. The test of the composite null hypothesis (3.3) consists of testing whether there is a $b \in \mathcal{B}$ for which the point $(\vartheta_0', b')'$ is contained in a confidence region for θ. Therefore, testing (3.3) can be reduced to testing (3.2). Define $\breve{\theta}(b) = (\vartheta_0', b')'$ for any $b \in \mathcal{B}$. Let $c_\alpha(b)$ denote the α-level critical value for testing the simple hypothesis $H_0: \theta = \breve{\theta}(b)$. That is, $c_\alpha(b)$ is the $1-\alpha$ quantile of the distribution of $T_n[\breve{\theta}(b)]$. If hypothesis (3.3) is correct, then the simple hypothesis $H_0: \theta = \breve{\theta}(\beta_0)$ is correct for some $\beta_0 \in \mathcal{B}$.

The critical values $c_\alpha(\theta_0)$ and $c_\alpha(b)$ are unknown in applications. Let $\hat{c}_\alpha(\theta_0)$ and $\hat{c}_\alpha(b)$, respectively, denote the consistent estimators of these quantities described in Section. 3.3. Hypothesis (3.2) is rejected at the α level if $T_n(\theta_0) > \hat{c}_\alpha(\theta_0)$. Hypothesis (3.3) is rejected at the α level if $T_n[\breve{\theta}(b)] > \hat{c}_\alpha(b)$ for every $b \in \mathcal{B}$. Computationally, the test of (3.3) consists of solving the nonlinear optimization problem

$$\text{(3.4)} \qquad \underset{b \in \mathcal{B}}{\text{minimize}} : \{T_n[\breve{\theta}(b)] - \hat{c}_\alpha(b)\}.$$

Hypothesis (3.3) is rejected if the optimal value of the objective function in (3.4) exceeds zero. Under hypothesis (3.3), the rejection probability does not exceed $P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)]$, where θ_0 is the true value of θ in (3.1). We obtain an upper bound on $P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)]$ that does not depend on θ_0. Therefore, it suffices to bound the probability of rejecting hypothesis (3.2).

In applications, the set $\mathcal{B}$ in (3.4) can be replaced by a much smaller set or even a single point if θ_0 is identfied. Let $\hat{\mathcal{B}}$ be an arbitrarily small neighborhood of $\hat{\theta}$. If H_0 is true, then $P(\theta_o \in \hat{\mathcal{B}}) \to 1$ at an exponentially fast rate as $n \to \infty$. Consequently, $\mathcal{B}$ in (3.4) can be replaced by an arbitrarily small neighborhood $\hat{\mathcal{B}}$. Moreover, the optimal value of the objective function of (3.4) is typically very close to $T_n(\hat{\theta}) - \hat{c}_\alpha(\hat{\theta})$. Therefore, $\breve{\theta}(b)$ can be replaced by $\hat{\theta}$ and $T_n[\breve{\theta}(b)] - \hat{c}_\alpha(b)$ with $T_n(\hat{\theta}) - \hat{c}_\alpha(\hat{\theta})$ in applications except, possibly, if $T_n(\hat{\theta}) - \hat{c}_\alpha(\hat{\theta})$ is close to 0. Solving the nonlinear optimization problem (3.4) is unnecessary if $T_n(\hat{\theta}) - \hat{c}_\alpha(\hat{\theta})$ is used in place of $T_n[\breve{\theta}(b)] - \hat{c}_\alpha(b)$.

The test based on $T_n(\theta_0)$ does not have optimal power in general, because $T_n(\theta_0)$ and its quantile analog that is described in Section 5 avoid estimation of the inverses of matrices that may be nearly singular. Estimates of inverses of nearly singular matrices can be very imprecise. Tests based on them can have low finite-sample power, and there can be large differences between the true and nominal probabilities with which such tests reject correct null hypotheses. Moreover, finite-sample results of the type presented here are difficult or impossible to obtain for the large class of models treated in this paper and statistics that depend on estimates of matrix inverses. The results of Spokoiny (2012) and Spokoiny and Zhilova (2015) illustrate this difficulty.

3.3 *The Test Procedure*

Under the simple null hypothesis (3.2),

(3.5) $$T_n(\theta_0) = n^{-1} \sum_{j=1}^{q} \left(\sum_{i=1}^{n} Z_{ij} U_i \right)^2 ,$$

where $U_i = Y_i - g(X_i, \theta_0)$. If the distribution of ZU were known, the finite-sample distribution of $T_n(\theta_0)$ could be computed from (3.5) by simulation. However, the distribution of ZU is unknown. To overcome this problem, define V to be the $q \times 1$ vector whose j'th component ($j = 1,...,q$) is

$$V_j = n^{-1/2} \sum_{i=1}^{n} Z_{ij} U_i .$$

Then $E(V) = 0$, $E(VV') = \Sigma$, and $T_n(\theta_0) = V'V$. Let $\hat{\Sigma}$ be a consistent estimator of Σ, and let $\hat{V}$ be a $q \times 1$ random vector that is distributed as $N(0, \hat{\Sigma})$. Define

(3.6) $$\hat{T}_n(\theta_0) = \hat{V}'\hat{V} .$$

The distribution of $\hat{T}_n(\theta_0)$ can be computed with any desired accuracy by simulation. Let $\hat{c}_\alpha(\theta_0)$ denote the $1-\alpha$ quantile of the distribution of $\hat{T}_n(\theta_0)$. Then

(3.7) $$P[\hat{T}_n(\theta_0) > \hat{c}_a(\theta_o)] = \alpha .$$

Section 4 presents a finite-sample upper bound on $| P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] - \alpha |$ that holds under H_0. Accordingly, the test procedure proposed here consists of:

1. Estimate Σ using the estimator $\hat{\Sigma}$ consisting of the $q \times q$ matrix whose (j,k) component is

$$\hat{\Sigma}_{jk} = n^{-1}\sum_{i=1}^{n} Z_{ij}Z_{ik}[Y_i - g(X_i,\theta_0)]^2 - \hat{\mu}_j\hat{\mu}_k ,$$

where

$$\hat{\mu}_j = n^{-1}\sum_{i=1}^{n} Z_{ij}[Y_i - g(X_i,\theta_0)] .$$

2. Use simulation to compute the distribution of $\hat{T}_n(\theta_0)$ and the critical value $\hat{c}_\alpha(\theta_0)$ by repeatedly drawing $\hat{V}$ from the $N(0,\hat{\Sigma})$ distribution.
3. Reject H_0 at the α level if $T_n(\theta_0) > \hat{c}_\alpha(\theta_0)$.

The critical value of $\hat{T}_n[\breve{\theta}(b)]$, $\hat{c}_\alpha(b)$, is estimated by replacing θ_0 with $\breve{\theta}(b)$ in steps 1-2. Section 5 presents Monte Carlo evidence on the numerical performance of this procedure.

It is not difficult to derive the asymptotic distribution of $T_n(\theta_0)$. See Theorem 4.2 in Section 4. This distribution depends on the unknown population parameter Σ. The finite-sample distribution of $\hat{T}_n(\theta_0)$ is the asymptotic distribution of $T_n(\theta_0)$ with Σ replaced by $\hat{\Sigma}$. Thus, the foregoing computational procedure is a simulation method to compute the estimated asymptotic distribution of $T_n(\theta_0)$. The main result for model (3.1), which is given in Theorem 4.1 and Corollary 4.1, is a bound on the difference between the unknown finite-sample probability that $T_n(\theta_0)$ rejects H_0 and the estimated asymptotic rejection probability. The latter is the finite-sample probability that $\hat{T}_n(\theta_0)$ rejects H_0. A similar result for the quantile version of $T_n(\theta_0)$ is given in Theorem 5.1. The distributions of $T_n(\theta_0)$, $\hat{T}_n(\theta_0)$, and their quantile versions are not chi-square because, to avoid the need for inverting estimated matrices, these statistics are not Studentized.

4. MAIN RESULT FOR MODEL (3.1)

4.1 *Inference under the Null* Hypothesis

This section presents the finite-sample upper bound on $|P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] - \alpha|$ in model (3.1) under the null hypothesis (3.2). As was explained in Section 3.2, testing hypothesis (3.3) can be reduced to testing hypothesis (3.2). Section 4.3 presents results on inference under alternative hypotheses.

Make the following assumptions, which are stated in a way that accommodates tests of both simple hypothesis (3.2) and composite hypothesis (3.3).

Assumption 1: $\{Y_i, X_i, Z_i :\ i = 1,...,n\}$ is an independent random sample from the distribution of (Y, X, Z). (ii) $\theta \in \Theta \subset \mathbb{R}^d$.

Assumption 2: (i) $\Sigma(\theta)$ is nonsingular for every $\theta \in \Theta$. (ii) Let $\Sigma_{jk}^{-1}(\theta)$ denote the (j,k) component of $\Sigma^{-1}(\theta)$. There is a constant $C_\Sigma < \infty$ such that $|\Sigma_{jk}^{-1}(\theta)| \le C_\Sigma$ for each $j,k = 1,...,q$ and every $\theta \in \Theta$.

Define the $q \times 1$ vectors $\xi = ZU$ and $\zeta = \Sigma^{-1/2}\xi$. Define the $q \times q$ matrix $\eta = ZZ'U^2$ Let ξ_j and ζ_j ($j = 1,...,q$) denote the j'th components of ξ and ζ, respectively. Let η_{jk} ($j,k = 1,...,q$) denote the (j,k) component of η.

Assumption 3: (i) There is a finite constant m_3 such that $E|\zeta_j|^3 \le m_3$ for every $j = 1,...,q$. (ii) There is a finite constant $\ell \ge \max[\max_j Var(\xi_j), \max_{j,k} Var(\eta_{jk})]$ such that $E|\xi_j|^r \le \ell^{r-1} r!$ and $E|\eta_{jk} - E(\eta_{jk})|^r \le \ell^{r-1} r!$ for every $r = 3,4,5,...$ and $j,k = 1,...,q$.

Assumption 1 specifies the sampling process. Assumption 2 establishes mild non-singularity conditions. For example, if U and Z are independent, then Assumption 2 requires $\text{cov}(Z)$ to be non-singular. Assumption 3 requires the distributions of the components of ξ and η to be thin-tailed. The assumption is satisfied, for example, if these distributions are sub-exponential. It is needed to establish the conditions of certain probability inequalities that are used in the proof of Theorem 4.1.

For any $t > 0$ define

$$r(t) = \left(\frac{6\ell t}{n}\right)^{1/2}$$

and

$$\tilde{r}(t) = C_\Sigma q^2 [r(t) + r(t)^2].$$

The following theorem and corollary give the finite-sample upper bound on $|P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] - \alpha|$ in model (3.1). It suffices to state the theorem for a test of hypothesis (3.2), because testing hypothesis (3.3) can be reduced to testing hypothesis (3.2).

Theorem 4.1: Let assumptions 1-3 and hypothesis (3.2) hold. Define $\hat{c}_\alpha(\theta_0)$ as in (3.7), but treat it as a non-stochastic constant in the following inequality. If $\max[\tilde{r}(t), r(t)] < 1$, then

$$(4.1) \qquad |P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] - \alpha| \le \frac{(42q^{1/4} + 16)q^{3/2} m_3}{n^{1/2}} + \min\begin{cases} q2^{q+1}\tilde{r}(t - 2\log q) \\ \frac{1}{\sqrt{2}}\{\tilde{r}(t - 2\log q) - \log[1 - \tilde{r}(t - 2\log q)]\}^{1/2} \end{cases}$$

with probability at least $1 - 4q^2 e^{-t}$. ■

The first term on the right-hand side of (4.1) bounds the error of the normal approximation to the distribution of $Z'U$. The second term bounds the error of approximating Σ by $\hat{\Sigma}$. The critical value $\hat{c}_\alpha(\theta_0)$ is a function of the random matrix $\hat{\Sigma}$ and, therefore, a random variable. However, in the probability expression on the left-hand side of (4.1), $\hat{c}_\alpha(\theta_0)$ is treated as a non-stochastic constant. When $\hat{c}_\alpha(\theta_0)$ is treated this way, inequality (4.1) holds only if $\hat{\Sigma}$ satisfies conditions (9.3) and (9.4) in the appendix. If these conditions are not satisfied, then (4.1) may not hold. The probability that the conditions are satisfied is at least $1-4q^2e^{-t}$.

The following corollary to Theorem 4.1 gives an unconditional bound on $|P[T_n(\theta_0)>\hat{c}_\alpha(\theta_0)]-\alpha|$. The corollary is a straightforward implication of Theorem 4.1 and, therefore, is presented without proof.

Corollary 4.1: Let assumptions 1-3 and hypothesis (3.2) hold. Define $\hat{c}_\alpha(\theta_0)$ as in (3.7) and

$$w_n(t)=\frac{(42q^{1/4}+16)q^{3/2}m_3}{n^{1/2}}+\min\begin{cases} q2^{q+1}\tilde{r}(t-2\log q) \\ \frac{1}{\sqrt{2}}\{\tilde{r}(t-2\log q)-\log[1-\tilde{r}(t-2\log q)]\}^{1/2}\end{cases}.$$

Then

(4.2) $$|P[T_n(\theta_0)>\hat{c}_\alpha(\theta_0)]-\alpha|\le \inf_{t:\max[\tilde{r}(t),r(t)]<1}\left[w_n(t)+4q^2e^{-t}\right]. \blacksquare$$

The right-hand sides of (4.1) and (4.2) do not depend on how X is related to the instruments. In particular, the upper bound on the probability of rejecting a correct simple or composite null hypothesis does not depend on the strength or weakness of the instruments.

The bounds on the right-hand sides of (4.1) and (4.2) decrease at the rate $n^{-1/2}$ as n increases if q remains fixed. If q increases as n increases, the bounds are $O(q^2/n^{1/2})$ and converge to zero if $q^4/n\to 0$. In practice, the left-hand sides of (4.1) and (4.2) are likely to be close to zero only if $q^2/n^{1/2}$ is close to zero. The ratio q^4/n is larger than the ratio obtained by several others. Newey and Windmeijer (2009) obtained asymptotic normality with $q^3/n\to 0$. Andrews and Stock (2007b) obtained a similar result for a linear simultaneous equations model. Faster rates of increase of q as a function of n are possible under stronger assumptions. See, for example, Bekker (1994). In contrast to these results, (4.1) and (4.2) are finite-sample inequalities, hold under weak distributional assumptions, and do not require linearity or simultaneous equations.

Like other large deviations bounds in statistics and the Berry-Esséen bound, the bounds on the right-hand sides of (4.1) and (4.2) and the bound given in Theorem 5.1 in Section 5 tend to be loose unless n is large because they accommodate "worst case" distributions of (Y, X, Z). For example, the distribution of $Z[Y - g(X, \theta_0)]$ may be far from multivariate normal or $\hat{\Sigma}$ may be an inaccurate estimate of Σ. The numerical performance of the test procedure of Section 3.3 in less extreme cases is illustrated in Section 6. Table 1 presents numerical illustrations of the bound of Corollary 4.2 in Section 4.2, which is a special case of the bound of Corollary 4.1.

In many applications of IV estimation, the model under consideration is linear with one endogenous covariate and one instrument. The bounds in (4.1) and (4.2) are much tighter in this case. Section 4.2 provides details. Tighter bounds also can be obtained, though at the cost of a reduction in power, by replacing $T_n(\theta)$ with the statistic

$$\hat{S}_n(\theta) = \max_{1 \le j \le q} \left| n^{-1/2} \sum_{i=1}^{n} \frac{Z_{ij}[Y_i - g(X_i, \theta_0)]}{\hat{\Sigma}_{jj}^{1/2}} \right| .$$

See Section 9.2 of the appendix.

The bounds in (4.1), (4.2), and Theorem 5.1 cannot be computed in applications because they depend on the unknown population parameters m_3 and ℓ. In some cases, however, an upper bound on m_3 can be obtained and knowledge of ℓ can be made unnecessary by restricting the thickness of the tails of the distribution of $Z'U$. Section 4.2 provides an example. Even if m_3 cannot be bounded and ℓ is unknown, the bounds in (4.1), (4.2), and Theorem 5.1 are useful for several reasons.

1. They show that it is possible to construct a test that is applicable to a wide variety of models and hypotheses and for which there is a finite-sample upper bound on the ERP.

2. They show, without relying on asymptotic approximations, that a simple, easily implemented test is valid uniformly over a wide class of distributions of (Y, X, Z), regardless of the strength or weakness of the instruments. The results in Section 5 show that uniform validity extends to linear and nonlinear quantile models.

3. They show how changing the sample size and number of instruments affects the maximum possible error in the test's rejection probability.

4.2 *Linear Models and the Frisch-Waugh-Lovell Transformation*

If $g(X, \theta)$ is a linear function of X, then (3.1) can be written as

$$Y = X\beta + W\gamma + U; \quad E(U \mid Z_X, W) = 0 ,$$

where Y is an $n \times 1$ vector; X is a $n \times q_X$ vector of endogenous variables; W is a $n \times q_W$ vector of exogenous variables; Z_X is a $n \times q_Z$ vector of instruments for X; β and γ, respectively, are $q_X \times 1$ and $q_W \times 1$ vectors of constant parameters; and $q_Z \ge q_X$. Define

$M = I - W(W'W)^{-1}W'$.

Then the Frisch-Waugh-Lovell transformation gives

$$MY = MX\beta + MU$$

and the moment condition

(4.3) $\quad EZ_X'M(Y - X\beta) = 0$.

The hypothesis H_0 : $\beta = \beta_0$ can be tested by applying the method described in Section 3 to the moment condition (4.3). The results stated in Theorem 4.1 and Corollary 4.1 apply after replacing q with q_Z .

Often in applications, $q_Z = 1$, so the test described in Section 3 requires only one instrument and moment condition, regardless of q_W . In this case, testing H_0 is equivalent to testing a simple hypothesis about a population mean. Non-Studentization of $T_n(\theta_0)$ does not cause a loss of asymptotic power. Moreover, the Bentkus (2003) and Raič (2019) multivariate Berry-Esséen inequality can be replaced by the univariate inequality in (4.2), thereby tightening the bound. The result is:

Corollary 4.2: Let the moment condition be (4.3) with $q_Z = 1$. Let the null hypothesis $H_0 : \beta = \beta_0$ and assumptions 1-3 hold. Define $\hat{c}_\alpha(\theta_0)$ as in (3.7) but with β_0 in place of θ_0. Define $\sigma_{FWL}^2 = Var[Z_X'M(Y - X\beta_0)]$, $\zeta_{FWL} = Z_X'M(Y - X\beta_0)/\sigma_{FWL}$, $m_{3,FWL} = E|\zeta_{FWL}|^3$, and

$$w_{n,FWL}(t) = \frac{0.9496 m_{3,FWL}}{n^{1/2}} + \min\begin{cases} 4\tilde{r}(t) \\ \dfrac{1}{\sqrt{2}}\{\tilde{r}(t) - \log[1 - \tilde{r}(t)]\}^{1/2} \end{cases}.$$

Then

$$|P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] - \alpha| \le \min_{t:\max[\tilde{r}(t), r(t)]<1} \left[w_{n,FWL}(t) + 4e^{-t} \right]. \blacksquare$$

Table 1 provides numerical illustrations of the bound in Corollary 4.2. As expected, the bound is large unless n is also large. This is due mainly to the bound on the random sampling error in estimating σ_{FWL}^2 (the min term in Corollary 4.2). If σ_{FWL}^2 is treated as known, the tabulated bounds are smaller by factors of roughly 5-10.

The bound can be tightened further and its dependence on ℓ removed if the distribution of $\breve{U} \equiv Z_X'M(Y - X\beta_0)$ is sub-Gaussian. The result is:

Corollary 4.3: Let the moment condition be (4.3) with $q_Z = 1$. Let the null hypothesis $H_0 : \beta = \beta_0$ and assumption 1 hold. Define $\hat{c}_\alpha(\theta_0)$ as in (3.7) but with β_0 in place of θ_0. Define σ_{FWL}^2

and $m_{3,FWL}$ as in Corollary 4.2. Assume that $0 < \sigma_{FWL}^2 < \infty$ and the distribution of $\breve{U}$ is sub-Gaussian with variance proxy σ_{SG}^2. That is

$$E\exp(s\breve{U}) \le \exp\left(\frac{\sigma_{SG}^2 s^2}{2}\right)$$

for all $s \in \mathbb{R}$. Let $\kappa_n = \min[n/2, 256\sigma_{SG}^2)$. Then

(4.4) $$|P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] - \alpha| \le 0.9496\frac{m_3}{n^{1/2}} + \min_{t:\le\kappa_n}\left[16\sqrt{2}\sigma_{SG}^2\left(\frac{t}{n}\right)^{1/2} + 3e^{-t}\right]. \blacksquare$$

Corollary 4.3 is proved in the appendix. The dependence of (4.4) on m_3 can be eliminated at the cost of a looser bound by using the inequality for sub-Gaussian random variables

$$m_3 \le 3\sqrt{2\pi}\sigma_{SG}^3.$$

The parameter σ_{SG}^2 characterizes the thickness of the tails of the distribution of $\breve{U}$. The dependence of (4.4) on this parameter cannot be eliminated.

4.3 *The Distribution of $T_n(\theta_0)$ under Alternative Hypotheses*

This section presents finite-sample bounds on the distribution of $T_n(\theta_0)$ under the alternative hypothesis

(4.5) $$H_1: \theta = \theta_0 + \kappa,$$

where κ is a finite $q \times 1$ vector. These bounds provide finite-sample lower and upper bounds on the power of the test of H_0 against H_1. This section also presents the asymptotic distribution of $T_n(\theta_0)$ under local alternative hypotheses with strong and weak instruments.

The following notation extends the notation of Theorem 4.1 to the case in which H_0 is false and H_1 is true. Define the $q \times q$ covariance matrix

$$\Sigma_H = Var\{Z'[Y - g(X,\theta_0)]\}$$

and the $q \times 1$ vector

$$\zeta_H = \Sigma_H^{-1/2} Z\{U + [g(X,\theta_0+\kappa) - g(X,\theta_0)]\} - E\{\Sigma_H^{-1/2} Z[g(X,\theta_0+\kappa) - g(X,\theta_0)]\}$$

if Σ_H is non-singular. Let ζ_{Hj} ($j = 1,...,q$) denote the j'th component of ζ_H. Define μ to be the $q \times 1$ vector whose j'th component is

$$\mu_j = E\{Z_{1j}[Y_1 - g(X_1,\theta_0]\} = E\{Z_{1j}[g(X_1,\theta_0+\kappa) - g(X_1,\theta_0)].$$

Let $\{\lambda_{Hj}: j=1,...,q\}$ denote the eigenvalues of Σ_H and Π denote the orthogonal matrix that diagonalizes Σ_H. That is $\Pi\Sigma_H\Pi' = \Lambda_H$, where Λ_H is the diagonal matrix whose diagonal elements are the eigenvalues of Σ_H. Define γ_{Hj} to be the j'th element of the $q \times 1$ vector

$$\gamma_H = n^{1/2}\Pi\Sigma_H^{-1/2}\mu .$$

Let $\{\chi_j^2(\gamma_{Hj}^2): j=1,...,q\}$ be independent random variables that are distributed as non-central chi-square with one degree of freedom and non-central parameters γ_{Hj}^2. Finally, define

$$w_{Hn} = \frac{(42q^{1/4}+16)q^{3/2}m_{H3}}{n^{1/2}}$$

We now have:

Theorem 4.2: Let assumptions 1-3 hold. Also assume that Σ_H is non-singular and there is a finite constant m_{H3} such that $E|\zeta_{Hj}|^3 \le m_{H3}$ for every $j=1,...,q$. Then

(4.6) $$P\left[\sum_{j=1}^{q}\lambda_{Hj}\chi_j^2(\gamma_{Hj}^2) > \hat{c}_\alpha(\theta_0)\right] - w_{Hn} \le P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)]$$

$$\le P\left[\sum_{j=1}^{q}\lambda_{Hj}\chi_j^2(\gamma_{Hj}^2) > \hat{c}_\alpha(\theta_0)\right] + w_{Hn}. \quad \blacksquare$$

The bounds in (4.6) are loose if n is small. The lower bound, which is a lower bound on the power of the $T_n(\theta_0)$ test, provides information about the relation between the asymptotic local power of the test and the strength or weakness of the instruments Z. To this end, let $\theta-\theta_0$ depend on n so that (4.5) becomes a sequence of local alternative hypotheses. Write this sequence as

(4.6) $$\theta = \theta_0 + n^{a-1/2}\kappa$$

for some a such that $0 \le a \le 1/2$. Let $\|.\|_E$ denote the Euclidean norm. Define

$$h(\theta) = E[Zg(X,\theta)] .$$

Make

Assumption 4: (i) $h(\theta)$ is twice continuously differentiable with respect to θ in a neighborhood of θ. (ii) There is a finite $b_1 > 0$ such that

$$\lim_{n\to\infty} n^a \left\|\frac{\partial h(\theta_0)}{\partial\theta}\kappa\right\|_E = b_1 .$$

(iii) For any sequence $\{\tilde{\theta}_n\}$ such that $\tilde{\theta}_n \to \theta_0$,

$$\lim_{n\to\infty} n^a \left\| \kappa' \frac{\partial^2 h(\tilde{\theta}_n)}{\partial\theta'\partial\theta} \kappa \right\|_E = \lim_{n\to\infty} n^a \left\| \kappa' \frac{\partial^2 h(\theta_0)}{\partial\theta'\partial\theta} \kappa \right\|_E = \begin{cases} 0 \text{ if } a < 1/2 \\ b_2 \text{ for some } b_2 < \infty \text{ if } a = 1/2. \end{cases}$$

The strength of the instrument Z is characterized by the value of a. Z is strong in the direction of κ if $a = 0$, semi-strong in the direction of κ if $0 < a < 1/2$, and weak in the direction of κ if $a = 1/2$. The definition of weakness includes the qualification "in the direction κ" because weak instruments can cause a test to have low power in some directions but not others. This characterization of the strength of instruments is consistent with the characterizations in terms of drifting sequences used by many others. See, for example, Cheng (2015).

Now let Assumption 4 hold. Assume that Σ_H is non-singular. Define $\gamma_{\kappa j}$ to be the j'th component of the $q \times 1$ vector

$$\gamma_\kappa = \begin{cases} \Pi\Sigma_H^{-1/2} \lim_{n\to\infty} n^a \frac{\partial h(\theta_0)}{\partial\theta}\kappa \text{ if } 0 \le a < 1/2 \\ \Pi\Sigma_H^{-1/2} \lim_{n\to\infty} n^{1/2} \left[\frac{\partial h(\theta_0)}{\partial\theta}\kappa + \frac{1}{2}\kappa' \frac{\partial^2 h(\theta_0)}{\partial\theta\partial\theta'}\kappa \right] \text{ if } a = 1/2. \end{cases}.$$

We then have

Corollary 4.4: Let assumptions 1-4 hold, and assume that Σ_H is non-singular. Let $\{\chi_j^2(\gamma_{\kappa j}^2): j = 1,...,q\}$ be independent random variables that are distributed as non-central chi-square with one degree of freedom and non-central parameters $\gamma_{\kappa j}^2$. Under the sequence of local alternative hypotheses (4.6)

$$T_n(\theta_0) \to^d \sum_{j=1}^d \lambda_{Hj} \chi_j^2(\gamma_{\kappa j}^2). \quad \blacksquare$$

It follows from Corollary 4.4 that if Z is strong in the direction κ, the α-level $T_n(\theta_0)$ test has power exceeding α against alternative hypotheses whose "distance" from θ_0 is $O(n^{-1/2})$. If Z is semi-strong or weak in the direction κ, then then the power of the α-level $T_n(\theta_0)$ test exceeds α against alternatives in the direction κ whose "distance" from θ_0 is $O(n^{a-1/2})$ for some $a \le 1/2$.

5. QUANTILE IV MODELS AND TESTING A PARAMETRIC MODEL AGAINST A NONPARAMETRIC ALTERNATIVE

Section 5.1 treats quantile IV models. Section 5.2 treats tests of model (3.1) and quantile IV models against a nonparametric alternative. Section 5.3 treats the asymptotic distributions of the test statistics under local alternative hypotheses.

5.1 *Inference in Quantile IV Models*

The quantile model is

(5.1) $Y = g(X,\theta) + U; \quad P(U \le 0 \mid Z) = a_Q$,

where $0 < a_Q < 1$. As in model (3.1), Y is the dependent variable, X is a possibly endogenous explanatory variable, and Z is an instrument for X. The null hypotheses to be tested are (3.2) and (3.3). However, as is explained in Section 3.2, testing hypothesis (3.3) can be reduced to testing hypothesis (3.2). Therefore, only a test of hypothesis (3.2) is described in this section. The test presented in this section is a modified version of the test presented in Sections 3 and 4. Thus, the same test applies to both mean and quantile IV models and can also be used to test a parametric mean or quantile IV model against a nonparametric alternative.

Let $\{Y_i, X_i, Z_i : i = 1,...,n\}$ be an independent random sample from the distribution of (Y, X, Z) in (5.1). Let Z_{ij} $(i = 1,...,n; j = 1,...,q\}$ denote the j'th component of Z_i. For any $\theta \in \Theta$, define

$$T_{Qn}(\theta) = n^{-1}\sum_{j=1}^{q}\left[\sum_{i=1}^{n} Z_{ij} W_{Qi}(\theta)\right]^2 ,$$

where

$$W_{Qi}(\theta) = I[Y_i - g(X_i,\theta) \le 0] - a_Q .$$

Define

$$W_Q(\theta) = I[Y - g(X,\theta) \le 0] - a_Q .$$

Denote the covariance matrix of the random vector $ZW(\theta)$ by $\Sigma_Q(\theta)$. Define $\Sigma_Q = \Sigma_Q(\theta_0)$, and let $\hat{\Sigma}_Q$ be the consistent estimator of Σ_Q that is defined in the next paragraph. The statistic for testing hypothesis (3.2) is $T_{Qn}(\theta_0)$. Let $\hat{V}_Q$ be a $q \times 1$ random vector that is distributed as $N(0, \hat{\Sigma}_Q)$. Define

(5.2) $\hat{T}_{Qn}(\theta_0) = \hat{V}_Q' \hat{V}_Q$.

Let $\hat{c}_{Q\alpha}(\theta_0)$ denote the $1-\alpha$ quantile of the distribution of $\hat{T}_{Qn}(\theta_0)$.

The test procedure is:

1. Estimate Σ_Q using the estimator $\hat{\Sigma}_Q$ consisting of the $q \times q$ matrix whose (j,k) component is

$$\hat{\Sigma}_{Qjk} = n^{-1}\sum_{i=1}^{n} Z_{ij} Z_{ik} W_{Qi}(\theta_0)^2 - \hat{\mu}_{Qj}\hat{\mu}_{Qk} ,$$

where

$$\hat{\mu}_{Qj} = n^{-1}\sum_{i=1}^{n} Z_{ij}W_{Qi}(\theta_0).$$

2. Use simulation to compute the distribution of $\hat{T}_{Qn}(\theta_0)$ and the critical value $\hat{c}_{Q\alpha}(\theta_0)$ by repeatedly drawing $\hat{V}_Q$ from the $N(0,\hat{\Sigma}_Q)$ distribution.
3. Reject H_0 at the α level if $T_{Qn}(\theta_0) > \hat{c}_{Q\alpha}(\theta_0)$

To obtain a finite-sample upper bound on $|P[T_{Qn}(\theta_0) > \hat{c}_{Q\alpha}(\theta_0)] - \alpha|$ make the following assumptions.

<u>Assumption Q1</u>: (i) $\{Y_i, X_i, Z_i : i = 1,...,n\}$ is an independent random sample from the distribution of (Y, X, Z). (ii) $\theta \in \Theta \subset \mathbb{R}^d$

<u>Assumption Q2</u>: (i) $\Sigma_Q(\theta)$ is nonsingular for every $\theta \in \Theta$. (ii) Let $\Sigma_{Qjk}^{-1}(\theta)$ denote the (j,k) component of $\Sigma_Q^{-1}(\theta)$. There is a constant $C_{Q\Sigma} < \infty$ such that $|\Sigma_{Qjk}^{-1}(\theta)| \le C_{Q\Sigma}$ for each $j,k = 1,...,q$ and every $\theta \in \Theta$.

Define the $q \times 1$ vectors $\xi_Q = ZW_Q(\theta_0)$ and $\zeta_Q = \Sigma_Q^{-1/2}\xi_Q$. Define the $q \times q$ matrix $\eta_Q = ZZ'W_Q(\theta_0)^2$ Let ξ_{Qj} and ζ_{Qj} ($j = 1,...,q$) denote the j'th components of ξ_Q and ζ_Q, respectively. Let η_{Qjk} ($j,k = 1,...,q$) denote the (j,k) component of η_Q.

<u>Assumption Q3</u>: (i) There is a finite constant m_{Q3} such that $E|\zeta_{Qj}|^3 \le m_{Q3}$ for every $j = 1,...,q$. (ii) There is a finite constant $\ell_Q \ge \max[\max_j E(\xi_{Qj}^2), \max_{j,k} E(\eta_{Qjk}^2)]$ such that $E|\xi_{Qj}|^r \le \ell_Q^{r-1} r!$ and $E|\eta_{Qjk}|^r \le \ell_Q^{r-1} r!$ for every $r = 3,4,5,...$ and $j,k = 1,...,q$.

For any $t > 0$ define

$$r_Q(t) = \left(\frac{6\ell_Q t}{n}\right)^{1/2}$$

and

$$\tilde{r}_Q(t) = C_{Q\Sigma} q^2 [r_Q(t) + r_Q(t)^2].$$

The following theorem gives the finite-sample upper bound on $|P[T_{Qn}(\theta_0) > \hat{c}_{Q\alpha}(\theta_0)] - \alpha|$.

<u>Theorem 5.1</u>: Let assumptions Q1-Q3 and hypothesis (3.2) hold. If $\max[\tilde{r}(t), r(t)] < 1$, then

(5.3) $$|P[T_{Qn}(\theta_0) > \hat{c}_{Q\alpha}(\theta_0)] - \alpha| \le \frac{(42q^{1/4}+16)q^{3/2}m_{Q3}}{n^{1/2}}$$

$$+\min\begin{cases} q2^{q+1}\tilde{r}_Q(t-2\log q) \\ \frac{1}{\sqrt{2}}\left\{\tilde{r}_Q(t-2\log q) - \log[1-\tilde{r}_Q(t-2\log q)]\right\}^{1/2} \end{cases}$$

with probability at least $1-4q^2e^{-t}$. ■

The treatment of the critical value $\hat{c}_{Q\alpha}(\theta_0)$ in (5.3) is the same as that of $\hat{c}_\alpha(\theta_0)$ in (4.1), and the interpretation of the probabilities in Theorem 5.1 is the same as in Theorem 4.1. The following corollary is analogous to Corollary 4.1.

Corollary 5.1: Let assumptions Q1-Q3 and hypothesis (3.2) hold. Define

$$w_{Qn}(t) = \frac{(42q^{1/4}+16)q^{3/2}m_3}{n^{1/2}} + \min\begin{cases} q2^{q+1}\tilde{r}_Q(t-2\log q) \\ \frac{1}{\sqrt{2}}\left\{\tilde{r}_Q(t-2\log q) - \log[1-\tilde{r}_Q(t-2\log q)]\right\}^{1/2} \end{cases}.$$

Then

$$|P[T_{Qn}(\theta_0) > \hat{c}_{Q\alpha}(\theta_0)] - \alpha| \le \inf_{t:\, \max[\tilde{r}(t), r(t)]<1} [w_{Qn}(t) + 4q^2e^{-t}].\ ■$$

The asymptotic distribution of $T_{Qn}(\theta_0)$ under the sequence of local alternative hypotheses (4.2) is given in Theorem 5.2 (iii) in Section 5.2.

5.2 *Testing a Parametric Model against a Nonparametric Alternative*

This section explains how the methods of Sections 3 and 5.1 can be used to carry out a test of a parametric mean or quantile IV model against a nonparametric alternative. As in Sections 3 and 5.1, the method presented in this section provides a finite-sample bound on the difference between the true and nominal probabilities of rejecting a correct null hypothesis.

Consider, first, model (3.1). Let G be a function that is identified by the relation

(5.4) $$E[Y - G(X) | Z] = 0,$$

where Y, X, and Z are as defined in Section 2.1. The null hypothesis, H_0^{NP}, tested in this section is

(5.5) $$G(x) = g(x,\theta)$$

for some $\theta \in \Theta$ and almost every $x \in \text{supp}(X)$, where g is a known function. The alternative hypothesis, H_1^{NP}, is that there is no $\theta \in \Theta$ such that (5.5) holds for almost every $x \in \text{supp}(X)$. Specifically, under H_1^{NP}, (5.6) $G(X) = g(X, \theta_0) + \Delta(X)$,

for some $\theta_0 \in \Theta$, where $\Delta(x)$ a function such that $E|Z_j \Delta(X)| < \infty$. To carry out the test, define $T_n(\theta)$ as in Section 3.1 and $\hat{c}_\alpha(\theta)$ as in Section 3.3 after replacing θ_0 with θ. The test of H_0^{NP} consists of solving the optimization problem

(5.7) $\underset{\theta \in \Theta}{\text{minimize}}: [T_n(\theta) - \hat{c}_\alpha(\theta)]$.

H_0^{NP} is rejected at the α level if the optimal value of the objective function in (5.7) exceeds zero. Theorem 4.1 and Corollary 4.1 provide finite-sample upper bounds on $|P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] - \alpha|$ under H_0 and, therefore, on the probability that a correct H_0 for model (3.1) is rejected at the α level.

Now consider model (5.1). The test of H_0^{NP} for model (5.1) consists of solving the optimization problem

$$\underset{\theta \in \Theta}{\text{minimize}}: [T_{Qn}(\theta) - \hat{c}_{Q\alpha}(\theta)].$$

Theorem 5.1 and Corollary 5.1 provide finite-sample upper bounds on $|P[T_{Qn}(\theta_0) > \hat{c}_{Q\alpha}(\theta_0)] - \alpha|$ under H_0^{NP} and, therefore, on the probability that a correct H_0^{NP} for model (5.1) is rejected at the α level.

5.3 *Asymptotic Distributions under Local Alternative Hypotheses*

It is not difficult to obtain finite-sample bounds, analogous to the bounds of Theorem 4.2, on the distributions of $T_n(\theta_0)$ and $T_{Qn}(\theta_0)$ under a nonparametric alternative to (5.1) or (5.5). Similarly, it is possible to obtain finite-sample bounds on the distribution of $T_{nQ}(\theta_0)$ under a parametric alternative to (5.1). As in Theorem 4.2 these bounds are loose if n is small but provide information about the asymptotic distributions of $T_n(\theta_0)$ and $T_{Qn}(\theta_0)$ under local alternative hypotheses. The remainder of this section presents asymptotic properties of $T_n(\theta_0)$ and $T_{Qn}(\theta_0)$ under local alternative hypotheses. The local alternatives are (4.6) for testing model 5.1 against a parametric local alternative hypothesis and

(5.8) $G(X) = g(X, \theta_0) + n^{a-1/2}\Delta(X)$

for testing models (3.1) and (5.1) against a nonparametric alternative hypothesis, where $0 \le a \le 1/2$ and $\Delta(X)$ satisfies Assumption Q4 below for model (3.1) and alternative hypothesis (5.8) or Q5 for model (5.1) and alternative hypothesis (5.8). Assumption Q6 below applies to model (5.1) and alternative hypothesis (4.6). These assumptions provide characterizations of the strength of instruments that are consistent with characterizations in terms of drifting sequences and extend the characterization of Assumption 4 to nonparametric alternatives and quantile models.

Assumption Q4 (Model 3.1 with a nonparametric alternative): (i) Alternative hypothesis (5.8) holds. (ii) There is a finite $b > 0$ such that $\lim_{n\to\infty} n^a \left\| E \mid Z_j \Delta(X) \mid \right\|_E = b$.

When assumption Q4 holds, define η_j^{NP} to be the j'th component of the $q \times 1$ vector

$$\eta^{NP} = \lim_{n\to\infty} n^a \Pi \Sigma^{-1/2} E[Z\Delta(X)].$$

Now consider model (5.1). Let $f_{U|X,Z}$ denote the probability density of U conditional on X, Z whenever this quantity exists. Let $\{\lambda_{Qj} : j = 1, 2, ..., q\}$ denote the eigenvalues of Σ_Q. Let Π_Q denote the orthogonal matrix that diagonalizes Σ_Q. For model (5.1) with alternative hypothesis (5.8), define

$$h_Q^{NP}(u) = -E[Zf_{U|X,Z}(u \mid X, Z)\Delta(X)]$$

and

$$\ell_Q^{NP}(u) = (1/2)E\{Zf_{U|X,Z}(u \mid X, Z)[\Delta(X)]^2\}$$

when $f_{U|X,Z}$ exists. Make

Assumption Q5 (Model 5.1 with a nonparametric alternative): (i) Alternative hypothesis (5.8) holds. (ii) $f_{U|X,Z}(u \mid x, z)$ exists and is a continuously differentiable function of u in a neighborhood of $u = 0$ for all (x, z). (iii) $\ell_Q^{NP}(u)$ is a continuously differentiable function of u for all u in a neighborhood of 0. (iv) There is a finite $b_1 > 0$ such that

$$\lim_{n\to\infty} n^a \left\| h_Q^{NP}(0) \right\|_E = b_1$$

(v) For any sequence $\{\tilde{u}_n\}$ such that $\tilde{u}_n \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} n^{2a-1/2} \left\| \frac{\partial \ell_Q^{NP}(\tilde{u}_n)}{\partial u} \right\|_E = 0.$$

When assumption Q5 holds, define η_{Qj}^{NP} to be the j'th component of the $q \times 1$ vector

$$\eta_Q^{NP} = \lim_{n\to\infty} n^a \Pi_Q \Sigma_Q^{-1/2} h_Q^{NP}(0).$$

Now consider model (5.1) and alternative hypothesis (4.6). Define

$$h_Q^P(\theta) = -n^{1/2} E\{Zf_{U|X,Z}(0 \mid X, Z)[g(X, \theta) - g(X, \theta_0)]\}$$

and

$$\ell_Q^P(u, \theta) = (1/2)n^{1/2} E\{Zf_{U|X,Z}(u \mid X, Z)[g(X, \theta) - g(X, \theta_0)]^2\}$$

when $f_{U|X,Z}$ exists. Make

Assumption Q6 (Model 5.1 with a parametric alternative): (i) Alternative hypothesis (4.6) holds. (ii) $f_{U|X,Z}(u \mid x,z)$ exists and is a continuously differentiable function of u in a neighborhood of $u=0$ for all (x,z). (iii) $h_Q^P(\theta)$ is twice continuously differentiable with respect to θ in a neighborhood of θ_0. (iv) $\ell_Q^P(u,\theta)$ is a continuously differentiable function of u for all u in a neighborhood of 0 and all θ in a neighborhood of θ_0. (v) There is a finite $b_1 > 0$ such that

$$\lim_{n\to\infty} n^a \left\| \frac{\partial h_Q^P(\theta_0)}{\partial \theta} \right\|_E = b_1 .$$

(v) For any sequences $\{\tilde{u}_n\}$ and $\{\tilde{\theta}_n\}$ such that $\tilde{u}_n \to 0$ and $n^{1/2-a}\left\|(\tilde{\theta}_n - \theta_0)\right\|_E$ is bounded as $n \to \infty$,

$$\lim_{n\to\infty} n^a \left\| \frac{\partial \ell_Q^{NP}(\tilde{u}_n, \tilde{\theta}_n)}{\partial u} \right\|_E = 0 .$$

When Assumption Q6 holds, define

$$\eta_Q^P = \lim_{n\to\infty} n^a \Pi_Q \Sigma_Q^{-1/2} h_Q^P(0) .$$

We now have

Theorem 5.2: (i) (Model 3.1 with a nonparametric alternative hypothesis). Let assumptions 1-3 and Q4 hold. Let $\{\chi_j^2[(\eta_j^{NP})^2]: j=1,...,q\}$ be independent random variables that are distributed as non-central chi-square with one degree of freedom and non-central parameters $(\eta_j^{NP})^2$. Under the sequence of local alternatives (5.8)

$$T_n(\theta_0) \to^d \sum_{j=1}^q \lambda_j \chi_j^2[(\eta_j^{NP})^2] .$$

(ii) (Model 5.1 with a nonparametric alternative hypothesis). Let assumptions Q1-Q3, and Q5 hold. Let $\{\chi_j^2[(\eta_{Qj}^{NP})^2]: j=1,...,q\}$ be independent random variables that are distributed as non-central chi-square with one degree of freedom and non-central parameters $(\eta_{Qj}^{NP})^2$. Under the sequence of local alternatives (5.8)

$$T_{Qn}(\theta_0) \to^d \sum_{j=1}^q \lambda_{Qj} \chi_j^2[(\eta_{Qj}^{NP})^2] .$$

(iii) (Model 5.1 with a parametric alternative hypothesis). Let assumptions Q1-Q3, and Q6 hold. Let $\{\chi_j^2[(\eta_{Qj}^P)^2]: j=1,...,q\}$ be independent random variables that are distributed as non-central chi-square with one degree of freedom and non-central parameters $(\eta_{Qj}^P)^2$. Under the sequence of local alternatives (4.6)

$$T_n(\theta_0) \to^d \sum_{j=1}^{q} \lambda_{Qj} \chi_j^2 [(\eta_{Qj}^P)^2] . \blacksquare$$

Theorem 5.2 implies that under its assumptions and with strong instruments, α-level tests based on $T_n(\theta_0)$ and $T_{Qn}(\theta_0)$ have asymptotic power exceeding α against parametric and nonparametric alternatives whose "distance" from H_0 is $O(n^{-1/2})$. With semi-strong or weak instruments, the tests have asymptotic power exceeding α against alternatives whose "distance" from H_0 is $O(n^{a-1/2})$.

6. MONTE CARLO EXPERIMENTS

This section reports the results of a Monte Carlo investigation of the numerical performance of the test procedure described in Section 3.2. Section 6.1 presents the results of experiments with a correct null hypothesis. Section 6.2 presents results about the power of the test.

6.1 *Probability of Rejecting a Correct Null Hypothesis*

The probability of rejecting the correct composite hypothesis (3.3) cannot exceed the probability of rejecting the correct simple hypothesis (3.2) with $\theta_0 = (\vartheta_0', \beta_0')$ for some β_0 such that θ_0 satisfies (3.1). Therefore, an upper bound on the probability of rejecting a correct simple or composite hypothesis can be obtained by carrying out an experiment with a simple hypothesis. Accordingly, experiments for correct null hypotheses were carried out only for simple hypotheses. When a simple hypothesis is correct,

$$T_n(\theta_0) = n^{-1} \sum_{j=1}^{q} \left[\sum_{i=1}^{n} Z_{ij} U_i \right]^2 .$$

The distribution of $T_n(\theta_0)$ does not depend on the function g or the distribution of X, so these are not specified in the designs of the experiments.

Experiments were carried out with sample sizes of $n = 100$ and $n = 1000$, and with $q = 1$, 2, 5, and 10 instruments. The instruments were sampled independently from the $N(0,1)$ distribution. Six distributions of U were used. These are:

1. The uniform distribution: $U \sim U[-2,2]$.

2. A mixture of the $N(0,1)$ and $N(2.5,1)$ distributions centered so that U has mean 0. The mixing probabilities are $p = 0.75$ and $p = 0.25$, respectively, for the $N(0,1)$ and $N(2.5,1)$ distributions. The resulting mixture distribution is skewed.

3. A mixture of the $N(0,1)$ and $N(4,1)$ distributions centered so that U has mean 0. The mixing probabilities are $p=0.75$ and $p=0.25$, respectively, for the $N(0,1)$ and $N(4,1)$ distributions. The resulting mixture distribution is bimodal.

4. The Laplace distribution with probability density function $f(u)=0.5e^{-|u|}$.

5. The Student t distribution with 10 degrees of freedom.

6. The difference between two independent lognormal distributions: $u=u_1-u_2$, where $\log u_j \sim N(0,1)$ for $j=1,2$.

Distributions 5 and 6 do not satisfy assumption 3. They illustrate the performance of T_n when assumption 3 is violated. The nominal rejection probability was 0.05. There were 10,000 Monte Carlo replications per experiment and 1000 Monte Carlo draws in the simulations used to compute critical values.

The results of the experiments are shown in Table 2. The differences between the empirical and nominal probabilities of rejecting H_0 are small when $q=1$. The empirical rejection probabilities tend to be below the nominal rejection probability of 0.05 when $q \geq 5$ and $n=100$. This is consistent with Theorem 4.1 in the sense that when $q^2/n^{1/2}$ increases, the bound on the difference between the true and nominal rejection probabilities increases. When $n=100$, $q^2/n^{1/2}=0.10$ if $q=1$, but $q^2/n^{1/2}=2.5$ if $q=5$ and 10 if $q=10$. The increases in the differences between the true and nominal rejection probabilities when $n=100$ reflect the large increases in the value of $q^2/n^{1/2}$ as q increases from 1 to 10.

6.2 *The Power of the Test*

This section presents Monte Carlo estimates of the power of the T_n test described in Section 3.2. To provide a basis for judging whether the power is high or low, the power of the T_n test is compared with the power of the S test of Stock and Wright (2000).

In the experiments reported in this section, data were generated from two models, one where H_0 is simple and one where it is composite. The model for the simple H_0 is

$$Y=\beta_0 X+U$$

$$X=\pi' Z+V$$

$$V=(1-\rho^2)^{1/2}\varepsilon+\rho U,$$

where $Z \sim N(0, I_q)$; I_q is the $q \times q$ identity matrix; U and ε have the distributions listed in Section 5.1; $\rho = 0.75$; $\beta_0 = 1.0$ or $\beta_0 = 0.20$, depending on the experiment; and $\pi = ce_q$, where e_q is a $q \times 1$ vector of ones and $c = 0.50$ or 0.25, depending on the experiment. The instruments are relatively strong when $c = 0.50$ and relatively weak when $c = 0.25$. The null hypothesis is H_0: $\beta_0 = 0$.

The model for the composite H_0 is

$$Y = \beta_1 X_1 + \beta_2 X_2 + U$$

$$X_1 = \pi' Z + V$$

$$V = (1 - \rho^2)^{1/2} \varepsilon + \rho U ,$$

where $Z \sim N(0, I_q)$; X_1 is the endogenous explanatory variable, X_2 is exogenous; X_2, U, and ε have the distributions listed in Section 6.1; $\rho = 0.75$; $\beta_1 = \beta_2 = 1$ or $\beta_1 = \beta_2 = 0.20$, depending on the experiment; and $\pi = ce_q$, where $c = 0.50$ or 0.25. The null hypothesis is H_0: $\beta_1 = 0$. With both models, the sample sizes are $n = 100$ and $n = 1000$, and the numbers of instruments are $q = 1$, 2, 5, and 10. The nominal level of the test is 0.05.

The results of the experiments with the simple H_0 are shown in Table 3 for $c = 0.50$ and Table 4 for $c = 0.25$. The results of the experiments with the composite H_0 are shown in Table 5 for $c = 0.50$ and Table 6 for $c = 0.25$. In most experiments, the power of the T_n test is similar to the power of the Stock-Wright (2000) test. This is not surprising because the T_n statistic is a non-Studentized version of the Stock-Wright (2000) statistic. However, the Stock-Wright (2000) test does not does not provide finite-sample bounds on the ERPs, and Stock and Wright (2000) do not investigate the relation between power and the degree of weakness of the instruments.

The power of the T_n test can be lower than the power of certain other tests if the number of instruments is large. However, the number of instruments is small (often one) in most applications. The power of the T_n test is similar to that of other tests when the number of instruments is small.

7. EMPIRICAL APPLICATIONS

This section presents two empirical applications of the T_n test. One consists of testing a hypothesis about a parameter in a finite-dimensional parametric model. The other consists of testing a parametric model against a nonparametric alternative.

7.1 *Testing a Hypothesis about a Parameter*

Acemoglu, Johnson, and Robinson (2001) (AJR) estimated models of the effect of institutions on economic performance. We consider the models in columns 1, 2, and 8 of Table IV of AJR. These models have the form

(7.1) $\log(GDP) = \beta_0 + \beta_1 AVPR + \gamma' X + \varepsilon; \quad E(\varepsilon \mid X, Z)$,

where GDP is a country's GDP per capita. $AVPR$ is an index of protection against expropriation risk. It is the institutional variable of interest and is potentially endogenous. X is a vector of exogenous covariates, Z is an instrument for $AVPR$, and ε is an unobserved random variable. The instrument Z is the logarithm of European settler mortality in a country. The β's and γ are constant parameters or vectors. The data consist of observations on 64 countries. The data and instrument are described in AJR. The parameter β_1 measures the effect of institutions on economic performance.

AJR present two stage least squares estimates of β_1 for the models in columns 1, 2, and 8 of Table IV and reject the hypothesis that $\beta_1 = 0$ in each column. We use T_n to test the hypothesis $\beta_1 = 0$ in each of the columns. The values of T_n are 50.987, 25.085, and 89864 for columns 1, 2, and 8, respectively. The corresponding 0.05-level critical values are 6.496, 9.461, and 10.845. The T_n test, like AJR, rejects the hypothesis that $\beta_1 = 0$ ($p < 0.05$).

7.2 *Testing a Parametric Model against a Nonparametric Alternative*

Blundell, Horowitz, and Parey (2012) (BHP) estimated parametric and non-parametric models of mean gasoline demand conditional on price and income. The parametric model is

(7.2) $\log Q = \beta_0 + \beta_1 \log P + \beta_2 \log Y + \varepsilon; \; E(\varepsilon \mid Y, Z) = 0$,

where Q is annual gasoline consumption by a household, P is the potentially endogenous price of gasoline, Y is the household's income, Z is an instrument for P, and ε is an unobserved random variable. The instrument Z is the distance between an oil platform in the Gulf of Mexico and the capital of the household's state. The β's are constant parameters. The nonparametric model is

(7.3) $\log Q = g(P, Y) + \varepsilon; \quad E(\varepsilon \mid Y, Z) = 0$,

where g is an unknown function. The data consist of 4812 observations from the 2001 National Household Travel Survey and are conditioned on a variety of demographic and geographical variables to reduce heterogeneity. BHP provide details about the data and explain the relevance and validity of the instrument.

Figure 2 of BHP shows graphs of the nonparametrically estimated demand function. The function appears nonlinear. We use T_n to test the hypotheses that the parametric model (7.2) is correctly specified against the nonparametric alternative (7.3). The value of T_n is $T_n = 61.21$. The 0.05-level critical value is 1.36. The T_n test rejects the hypothesis that the demand function is linear $(p < 0.05)$.

8. CONCLUSIONS

This paper has presented a method for carrying out inference in a wide variety of linear and nonlinear models estimated by instrumental variables. The results include finite-sample bounds on the differences between the true and nominal rejection probabilities of hypothesis tests based on the method. The test statistic is a non-Studentized version of the S test of Stock and Wright (2000) but is implemented and analyzed differently. The method is easy to implement and does not require strong distributional assumptions, linearity of the estimated model, or simultaneous equations. Nor does it require knowledge of the strength of the instruments or identification of the parameter about which inference is made. The method can be applied to quantile IV models that may be nonlinear and can be used to test a parametric IV or quantile IV model against a nonparametric alternative. The results presented here hold in finite samples, regardless of the strength of the instruments. The results of Monte Carlo experiments and two empirical applications have illustrated the numerical performance of the method.

9. APPENDIX: PROOFS OF THEOREMS

9.1. *Proofs of Theorems 3.1, 3.2, 5.1, and 5.2 and of Corollaries 4.3 and 4.4.*

Assumptions 1-3 and hypothesis (3.2) hold for lemmas 9.1-9.3.

Lemma 9.1: Let $\{v_i : i = 1,...,n\}$ be independent random $q \times 1$ vectors with the $N(0, I_{q\times q})$ distribution. Define

$$\tilde{T}_n(\theta_0) = \left(n^{-1/2} \sum_{i=1}^{n} v_i' \right) \Sigma \left(n^{-1/2} \sum_{i=1}^{n} v_i \right).$$

Then

$$\sup_{a \geq 0} | P[T_n(\theta_0) \leq a] - P[\tilde{T}_n(\theta_0) \leq a] | \leq \frac{(42q^{1/4} + 16)q^{3/2} m_3}{n^{1/2}}. \tag{9.1}$$

Proof: For each $i = 1,...,n$, define

$$\tilde{V}_i = \Sigma^{-1/2}(Z_i U_i).$$

Then $E(\tilde{V}_i) = 0$, $E(\tilde{V}_i \tilde{V}_i') = I_{q\times q}$, and

$$T_n(\theta_0) = \left(n^{-1/2} \sum_{i=1}^{n} \tilde{V}_i \right)' \Sigma \left(n^{-1/2} \sum_{i=1}^{n} \tilde{V}_i \right).$$

For any $a \geq 0$, the set

$$A = \{\tilde{V}_1, ..., \tilde{V}_n : \; T_n(\theta_0) \leq a\}$$

is convex. Therefore, (8.1) follows from Theorem 1.1 of Bentkus (2003). See, also, Corollary 11.1 of Dasgupta (2008). Q.E.D.

Define $r(t)$ as in Theorem 3.1. Define $\omega = \hat{\Sigma} - \Sigma$.

Lemma 9.2: For any $t > 0$ such that

(9.2) $\quad r(t) \leq 1$,

(9.3) $\quad |\,\omega_{jk}\,| \leq r(t) + r(t)^2$

and

(9.4) $\quad |\,(\Sigma^{-1}\omega)_{jk}\,| \leq C_\Sigma q[r(t) + r(t)^2]$

uniformly over $j,k = 1,...,q$ with probability at least $1 - 4q^2 e^{-t}$.

Proof: Define

$$\mu_j = EZ_{1j}[Y_1 - g(X_1, \theta_0)].$$

Then

$$|\,\omega_{jk}\,| = n^{-1} \left| \sum_{i=1}^{n} [Z_{ij} Z_{ik} U_i^2 - E(Z_{ij} Z_{ik} U_i^2)] - (\hat{\mu}_j - \mu_j)(\hat{\mu}_k - \mu_k) \right|$$

$$\leq n^{-1} \left| \sum_{i=1}^{n} [Z_{ij} Z_{ik} U_i^2 - E(Z_{ij} Z_{ik} U_i^2)] \right| + |\,(\hat{\mu}_j - \mu_j)(\hat{\mu}_k - \mu_k)\,|.$$

Bernstein's inequality gives

$$P\left[n^{-1} \left| \sum_{i=1}^{n} [Z_{ij} Z_{ik} U_i^2 - E(Z_{ij} Z_{ik} U_i^2)] \right| \geq r(t) \right] \leq 2e^{-t}$$

for each $(j,k) = 1,...,q$ and

(9.5) $\quad P[\,|\,\hat{\mu}_j - \mu_j\,| \geq r(t)] \leq 2e^{-t}$

for each $j = 1,...,q$. Therefore,

$$P\left[\max_{j,k} |\,\omega_{jk}\,| < r(t) + r(t)^2 \right] > 1 - 4q^2 e^{-t},$$

thereby establishing (9.3). In addition, (9.6) $\quad |(\Sigma^{-1}\omega)_{jk}| \le C_\Sigma \sum_{\ell=1}^{q} |\omega_{\ell k}|$.

Therefore, inequality (9.4) follows from (9.3) and (9.6). Q.E.D.

Define the random variables $V \sim N(0,\Sigma)$ and, conditional on $\hat{\Sigma}$, $\hat{V} \sim N(0,\hat{\Sigma})$. Also define

$$\Xi_n = \sup_a |P[\tilde{T}_n(\theta_0) \le a] - P[\hat{T}_n(\theta_0) \le a]| = \sup_a |P(V'V \le a) - P(\hat{V}'\hat{V} \le a)|.$$

Lemma 9.3: Define $\tilde{r}(t)$ as in Theorem 3.1. For any $t > 0$ such that (9.2) holds and $\tilde{r}(t) < 1$,

$$\Xi_n \le \min \begin{cases} q2^{q+1}\tilde{r}(t)] \\ \dfrac{1}{\sqrt{2}}\{\tilde{r}(t) - \log[1-\tilde{r}(t)]\}^{1/2} \end{cases}.$$

with probability at least $1 - 4q^2 e^{-t}$.

Proof: Let $TV(P_1, P_2)$ be the total variation distance between distributions P_1 and P_2. For any set $\mathcal{S} \subset \mathbb{R}^q$ and $q \times 1$ random vector v , define and $\mathcal{S}_\Sigma = \{v : \Sigma^{1/2} v \in \mathcal{S}\}$. Then,

$$P(\hat{V} \in \mathcal{S}) - P(V \in \mathcal{S}) = P(\Sigma^{-1/2}\hat{V} \in \mathcal{S}_\Sigma) - P(\Sigma^{-1/2}V \in \mathcal{S}_\Sigma).$$

By the definition of the total variation distance,

$$\Xi_n \le \sup_{\mathcal{S}} |P(\Sigma^{-1/2}\hat{V} \in \mathcal{S}_\Sigma) - P(\Sigma^{-1/2}V \in \mathcal{S}_\Sigma)| \le TV[N(0, I_{q\times q}), N(0, \Sigma^{-1}\hat{\Sigma})],$$

By DasGupta (2008, p. 23),

$$TV[N(0, I_{p\times p}), N(0, \Sigma^{-1}\hat{\Sigma})] \le \min \begin{cases} q2^{q+1}\left\|\Sigma^{-1}\hat{\Sigma} - I_{q\times q}\right\| \\ \dfrac{1}{\sqrt{2}}\left[Tr(\Sigma^{-1}\hat{\Sigma} - I_{q\times q}) - \log\det(\Sigma^{-1}\hat{\Sigma})\right]^{1/2} \end{cases},$$

where for any $q \times q$ matrix A,

$$\|A\|^2 = \sum_{j,k=1}^{q} a_{jk}^2 .$$

But

$$\Sigma^{-1}\hat{\Sigma} - I_{q\times q} = \Sigma^{-1}(\hat{\Sigma} - \Sigma) = \Sigma^{-1}\omega,$$

$$|(\Sigma^{-1}\omega)_{jk}| \le \sum_{\ell=1}^{q} |\Sigma_{j\ell}^{-1}\omega_{\ell k}| \le C_\Sigma \sum_{\ell=1}^{q} |\omega_{\ell k}|,$$

and

$$\left\|\Sigma^{-1}\hat{\Sigma} - I_{q\times q}\right\| \le C_\Sigma q^{1/2}\left[\sum_{k=1}^{q}\left(\sum_{\ell=1}^{q}|\omega_{\ell k}|\right)^2\right]^{1/2} \le C_\Sigma q^2 \max_{\ell,k} |\omega_{\ell k}|.$$

By Lemma 9.2

$$P\left[\max_{j,k} |\omega_{jk}| < r(t) + r(t)^2\right] > 1 - 4q^2 e^{-t}.$$

Therefore,

$$q2^{q+1}\left\|\Sigma^{-1}\hat{\Sigma} - I_{q\times q}\right\| \le q2^{q+1}\tilde{r}(t)$$

with probability exceeding $1 - 4q^2 e^{-t}$.

Now consider

$$\frac{1}{\sqrt{2}}\left[Tr(\Sigma^{-1}\hat{\Sigma} - I_{q\times q}) - \log\det(\Sigma^{-1}\hat{\Sigma})\right]^{1/2}.$$

We have

$$Tr(\Sigma^{-1}\hat{\Sigma} - I_{q\times q}) = Tr(\Sigma^{-1}\omega).$$

But

$$(\Sigma^{-1}\omega)_{jj} \le C_\Sigma q[r(t) + r(t)^2]$$

with probability exceeding $1 - 4q^2 e^{-t}$. Therefore

(9.7) $\quad Tr(\Sigma^{-1}\omega) \le \tilde{r}(t)$,

and

$$\frac{1}{\sqrt{2}}\left[Tr(\Sigma^{-1}\hat{\Sigma} - I_{q\times q}) - \log\det(\Sigma^{-1}\hat{\Sigma})\right]^{1/2} \le \frac{1}{\sqrt{2}}\left[\tilde{r}(t) - \log\det(\Sigma^{-1}\hat{\Sigma})\right]^{1/2}$$

with probability exceeding $1 - 4q^2 e^{-t}$.

In addition,

$$\log\det(\Sigma^{-1}\hat{\Sigma}) = \log\det(I_{q\times q} + \Sigma^{-1}\omega).$$

Let $\tilde{r}(t) < 1$. By Corollary 1 of Brent, Osborne, and Smith (2015)

(9.8) $\quad \det(I_{q\times q} + \Sigma^{-1}\omega) \ge 1 - \tilde{r}(t)$

and

$$\log\det(I_{q\times q} + \Sigma^{-1}\omega) \ge \log[1 - \tilde{r}(t)]$$

with probability exceeding $1 - 4q^2 e^{-t}$. Therefore,

$$\frac{1}{\sqrt{2}}\left[Tr(\Sigma^{-1}\hat{\Sigma} - I_{p\times p}) - \log\det(\Sigma^{-1}\hat{\Sigma})\right]^{1/2} \le \frac{1}{\sqrt{2}}\{\tilde{r}(t) - \log[1 - \tilde{r}(t)]\}^{1/2}$$

and

$$\Xi_n \le \min \begin{cases} C_\Sigma q^3 2^{q+1} \tilde{r}(t)] \\ \dfrac{1}{\sqrt{2}} \{\tilde{r}(t) - \log[1-\tilde{r}(t)]\}^{1/2} \end{cases}$$

with probability at least $1-4q^2 e^{-t}$. Q.E.D.

Proof of Theorem 4.1: Define $\tilde{T}_n(\theta_0)$ as in Lemma 9.1. By the triangle inequality

$$\sup_{a\ge 0} | P[T_n(\theta_0) \le a] - P[\hat{T}_n(\theta_0) \le a] |$$

$$= \sup_{a\ge 0} | P[T_n(\theta_0) \le a] - P[\tilde{T}_n(\theta_0) \le a] + P[\tilde{T}_n(\theta_0) \le a] - P[\hat{T}_n(\theta_0) \le a] |$$

$$\le \sup_{a\ge 0} \{| P[T_n(\theta_0) \le a] - P[\tilde{T}_n(\theta_0) \le a] | + | P[\tilde{T}_n(\theta_0) \le a] - P[\hat{T}_n(\theta_0) \le a] |\} \quad .$$

$$\le \sup_{a\ge 0} | P[T_n(\theta_0) \le a] - P[\tilde{T}_n(\theta_0) \le a] | + \sup_{a\ge 0} | P[\tilde{T}_n(\theta_0) \le a] - P[\hat{T}_n(\theta_0) \le a] |$$

$$\le \sup_{a\ge 0} | P[T_n(\theta_0) \le a] - P[\tilde{T}_n(\theta_0) \le a] | + \Xi_n .$$

Now combine lemmas 9.1 and 9.3. Q.E.D.

Proof of Corollary 4.3: The proof takes place in four steps.

Step 1 (Modification of Lemma 9.1): Replace the Bentkus (2002) inequality with the Berry-Esséen inequality with the constant of Shevtsova (2011) to obtain

(9.9) $$\sup_{a\ge 0} | P[T_n(\theta_0) \le a] - P[\tilde{T}_n(\theta_0) \le a] | \le 0.9496 \frac{m_3}{n^{1/2}} .$$

Step 2 (Modification of Lemma 9.2): Let H_0 be true. Define $\xi = Z_X' M(Y - X\beta_0)$. Then $E(\xi) = 0$. Define $\sigma_\xi^2 = E(\xi^2)$. Let ξ_i ($i = 1,...,n$) be i'th component of ξ. Define

$$\hat{\mu} = n^{-1} \sum_{i=1}^n \xi_i ,$$

$$\hat{\sigma}_\xi^2 = n^{-1} \sum_{i=1}^n (\xi_i^2 - \hat{\mu}^2) ,$$

and $\hat{\omega} = \hat{\sigma}_\xi^2 - \sigma_\xi^2$. Then

$$|\hat{\omega}| \le n^{-1} \left| \sum_{i=1}^n (\xi_i^2 - \sigma_\xi^2) \right| + \hat{\mu}^2$$

Bernstein's inequality for sub-Gaussian random variables yields

$$P(\hat{\mu} > t) \le 2\exp\left(-\frac{nt^2}{2\sigma_{SG}^2}\right)$$

for any $t > 0$ and, therefore,

$$P(\hat{\mu}^2 > t) \le 2\exp\left(-\frac{nt}{2\sigma_{SG}^2}\right). \quad (9.10)$$

Because ξ is sub-Gaussian, $\xi^2 - E(\xi^2)$ is sub-exponential. Bernstein's inequality for sub-exponential random variables yields

$$P\left[\left|n^{-1}\sum_{i=1}^{n}(\xi_i^2 - \sigma_\xi^2)\right| > t\right] \le \exp\left\{-\frac{n}{2}\left[\frac{t^2}{(16\sigma_{SG}^2)^2} \wedge \frac{t}{16\sigma_{SG}^2}\right]\right\} \quad (9.11)$$

for any $t > 0$. Combining (9.10) and (9.11) yields

$$P(|\hat{\omega}| > t) \le \exp\left\{-\frac{n}{2}\left[\frac{t^2}{(16\sigma_{SG}^2)^2} \wedge \frac{t}{16\sigma_{SG}^2}\right]\right\} + 2\exp\left(-\frac{nt}{2\sigma_{SG}^2}\right)$$

for any $t > 0$.

Step 3 (Modification of Lemma 9.3): As in the proof of Lemma 9.3,

$$\sup_t |P(\hat{V} \le t) - P(V \le t)| \le 4\frac{|\hat{\sigma}_\xi^2 - \sigma_\xi^2|}{\sigma_\xi^2}$$

Moreover,

$$P\left(|\hat{\sigma}_\xi^2 - \sigma_\xi^2| > \frac{\sigma_\xi^2 \tau}{4}\right) = P\left(|\hat{\omega}| > \frac{\sigma_\xi^2 \tau}{4}\right).$$

for any $\tau > 0$. Define t by

$$\frac{\sigma_\xi^2 \tau}{4} = 16\sqrt{2}\sigma_{SG}^2\left(\frac{t}{n}\right)^{1/2},$$

and assume that

$$t \le \min(n/2, 256\sigma_{SG}^2) \equiv \kappa_n .$$

Then it follows from Bernstein's inequality for sub-exponential random variables that

$$P\left[4\frac{|\hat{\sigma}_\xi^2 - \sigma_\xi^2|}{\sigma_\xi^2} > 16\sqrt{2}\sigma_{SG}^2\left(\frac{t}{n}\right)^{1/2}\right] \le 3e^{-t} .$$

Step 4 (Modification of proof of Theorem 4.1): Replace the results of Lemmas 9.1-9.3 with those of steps 1-3 in the proof or Theorem 4.1 to obtain

$$|P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] - \alpha| \le 0.9496\frac{m_3}{n^{1/2}} + 16\sqrt{2}\sigma_{SG}^2\left(\frac{t}{n}\right)^{1/2}$$

with probability at least $1-3e^{-t}$ and, therefore,

$$|P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] - \alpha| \le 0.9496\frac{m_3}{n^{1/2}} + \min_{t \le \kappa_n}\left[16\sqrt{2}\sigma_{SG}^2\left(\frac{t}{n}\right)^{1/2} + 3e^{-t}\right].$$ Q.E.D.

Proof of Theorem 4.2: Define $\tilde{V}_i$ as the $q\times 1$ vector whose j'th component is

$$\tilde{V}_{ij} = \Sigma_H^{-1/2}\{Z_{ij}[Y_i - g(X_i,\theta_0)] - \mu_j\}$$

$$= \Sigma_H^{-1/2}\{Z_{ij}U_i + Z_{ij}[g(X_i,\theta_0+\kappa) - g(X_i,\theta_0)] - \mu_j\}.$$

Then $E\tilde{V}_i = 0$ and $E(\tilde{V}_i\tilde{V}_i') = I_{q\times q}$. Under H_1,

$$T_n(\theta_0) = \left[n^{-1/2}\sum_{i=1}^{n}(\tilde{V}_i + \Sigma_H^{-1/2}\mu)\right]'\Sigma_H\left[n^{-1/2}\sum_{i=1}^{n}(\tilde{V}_i + \Sigma_H^{-1/2}\mu)\right]$$

Let $\{\nu_i : i = 1,...,n\}$ be independent random vectors with the $N(0, I_{q\times q})$ distribution. Define

$$\tilde{T}_n = \left[n^{-1/2}\sum_{i=1}^{n}(\nu_i + \Sigma_H^{-1/2}\mu)\right]'\Sigma_H\left[n^{-1/2}\sum_{i=1}^{n}(\nu_i + \Sigma_H^{-1/2}\mu)\right].$$

Then

(9.12) $$\tilde{T}_n \sim \sum_{j=1}^{q}\lambda_{Hj}\chi_j^2(\gamma_{Hj}^2).$$

Arguments like those used to prove Lemma 9.1 show that

(9.13) $$\sup_{a\ge 0}|P[T_n(\theta_0) \le a] - P(\tilde{T}_n \le a)| \le w_{Hn}.$$

Let $a = \hat{c}_\alpha(\theta_0)$ Then

$$P[\tilde{T}_n \le \hat{c}_\alpha(\theta_0)] - w_{Hn} \le P[T_n(\theta_0) \le \hat{c}_\alpha(\theta_0)] \le P[\tilde{T}_n \le \hat{c}_\alpha(\theta_0)] + w_{Hn}$$

and

$$P[\tilde{T}_n > \hat{c}_\alpha(\theta_0)] - w_{Hn} \le P[T_n(\theta_0) > \hat{c}_\alpha(\theta_0)] \le -P[\tilde{T}_n > \hat{c}_\alpha(\theta_0)] + w_{Hn}.$$

Q.E.D.

Proof of Corollary 4.4: As $n\to\infty$, $w_{Hn}\to 0$ and $\gamma_H \to \gamma_\kappa$. The corollary now follows from (9.12) and (9.13). Q.E.D.

Let F_{XZ} denote the distribution function of (X,Z) and $F_{X|Z}$ denote the conditional distribution function of X given Z.

Proof of Theorem 5.2:

Part (i): Part (i) follows from the multivariate generalization of the Lindeberg-Lévy central limit theorem and the definition of η^{NP}.

Part (ii): It follows from the multivariate generalization of the Lindeberg-Levy central limit theorem that

$$n^{-1/2}\sum_{i=1}^{n} Z_{ij}\{I[Y_i - g(X_i,\theta_0)\le 0]-a_Q\} \to^d \xi_{Qj}+\tilde{\tau}_{Qj},$$

where ξ_{Qj} is the j'th component of the $q\times 1$ random vector $\xi_Q \sim N(0,\Sigma_Q)$ and

$$\tilde{\tau}_{Qj} = \lim_{n\to\infty} n^{1/2} E\left(Z_{1j}\{I[U+n^{a-1/2}\Delta(X)\le 0]-a_Q\}\right)$$

$$= \lim_{n\to\infty} n^{1/2} EZ_{1j}\{F_{U|X,Z}[-n^{a-1/2}\Delta(X)\,|\,X_1,Z_1]-a_Q\}.$$

Define $\tilde{\tau}_Q$ as the $q\times 1$ vector whose j'th component is $\tilde{\tau}_{Qj}$. Then,

$$\tilde{\tau}_Q = \lim_{n\to\infty} n^{1/2}\int z\left\{F_{U|X,Z}[-n^{a-1/2}\Delta(x)\,|\,x,z]dF_{XZ}(x,z)-a_Q\right\}$$

$$= \lim_{n\to\infty} n^{1/2} EZ\{F_{U|X,Z}[-n^{a-1/2}\Delta(x)\,|\,X,Z]-a_Q\}.$$

By a Taylor series expansion and Assumption Q5(v),

$$\tilde{\tau}_Q = \lim_{n\to\infty} n^a\left[h_Q^{NP}(0)+n^{a-1/2}\frac{\partial \ell_Q^{NP}(0)}{\partial u}\right] = \lim_{n\to\infty} n^a h_Q^{NP}(0).$$

Therefore,

$$\eta_Q^{NP} = -\lim_{n\to\infty} n^a \Pi\Sigma^{-1/2} h_Q^{NP}(0).$$

Part (ii) now follows from the properties of quadratic forms of normally distributed random vectors.

Part (iii): Under local alternative hypothesis (4.6),

$$h(\theta) = n^{a-1/2}\frac{\partial h(\theta_0)}{\partial\theta'}\kappa + \frac{1}{2}n^{2a-1}\kappa'\frac{\partial^2 h(\tilde{\theta})}{\partial\theta\partial\theta'}\kappa,$$

where $\tilde{\theta}$ is between θ_0 and $\theta_0+n^{a-1/2}\kappa$. Therefore, the arguments made for Part (ii) apply to local alternative hypothesis (4.6) after replacing $\Delta(x)$ with $g(X,\theta_0+n^{a-1/2}\kappa)-g(X,\theta_0)$. It follows that under local alternative hypothesis (4.6)

$$T_n(\theta_0) \to^d \sum_{j=1}^{q}\lambda_{Qj}\chi_j^2[(\eta_{Qj}^P)^2].$$

This proves Theorem 5.2(iii). Q.E.D.

9.2 *The Statistic* $\hat{S}_n$

Assume throughout this section that model (3.1) null hypothesis (3.2) hold. Similar results can be obtained for model (5.1).

Define $\sigma_j = \Sigma_{jj}^{1/2}$, $\hat{\sigma}_j = \hat{\Sigma}_{jj}^{1/2}$,

$$S_n = \max_j \left| n^{-1/2} \sum_{i=1}^{n} \frac{Z_{ij}[Y_i - g(X_i, \theta_0)]}{\sigma_j} \right|$$

and

$$\hat{S}_n = \max_j \left| n^{-1/2} \sum_{i=1}^{n} \frac{Z_{ij}[Y_i - g(X_i, \theta_0)]}{\hat{\sigma}_j} \right|.$$

Let V and $\hat{V}$ be $q \times 1$ random vectors that are distributed as $N(0, \Sigma)$ and $N(0, \hat{\Sigma})$, respectively. Let V_j and $\hat{V}_j$, respectively, denote the j'th components of V and $\hat{V}$. Let $\hat{c}_\alpha$ be the $1-\alpha$ quantile of the distribution of $\sup_j (|\hat{V}_j| / \hat{\sigma}_j)$ with $\hat{\Sigma}$ treated as non-stochastic. Observe that $\hat{c}_\alpha$ can be estimated with arbitrary accuracy by simulation. This section gives lower and upper bounds on

$$P(\hat{S}_n \le t) - P\left(\sup_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t \right)$$

uniformly over t and, therefore, on

$$P(\hat{S}_n \le \hat{c}_\alpha) - P\left(\sup_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le \hat{c}_\alpha \right).$$

Define $m_{3j} = E|Z_{1j}[Y_1 - g(X_1, \theta_0)]|^3 / \sigma_j^3$ and $m_3 = \max_j m_{3j}$. Also define

$$S_{nj} = n^{-1/2} \sum_{i=1}^{n} \frac{Z_{ij}[Y_i - g(X_i, \theta_0)]}{\sigma_j}$$

and

$$\hat{S}_{nj} = n^{-1/2} \sum_{i=1}^{n} \frac{Z_{ij}[Y_i - g(X_i, \theta_0)]}{\hat{\sigma}_j}.$$

Then

$$\hat{S}_n = \max_j |\hat{S}_{nj}|.$$

For any $t > 0$ define $\mathcal{S}_t = [-t, t]^q$.

Lemma 9.4: Let assumption 1 hold, and let $0 < \sigma_j < \infty$ for all $j = 1,...,q$. Treat $\hat{\sigma}_j$ as non-stochastic. Then uniformly over $-\infty < t < \infty$

$$1 - \sum_{j=1}^{q} P(|V_j| > \hat{\sigma}_j t) - \frac{qm_3}{n^{1/2}} \le P(\hat{S}_n \in \mathcal{S}_t) \le \sum_{j=1}^{q} P(|V_j| \le \hat{\sigma}_j t) + \frac{qm_3}{n^{1/2}}.$$

Proof: The proof takes place in two steps.

Step 1: The lower bound.

$$P(\hat{S}_n \notin \mathcal{S}_t) = P\left[\bigcup_{j=1}^{q} (|\hat{S}_{nj}| > t)\right]$$

$$\le \sum_{j=1}^{q} P(|\hat{S}_{nj}| > t).$$

Therefore,

$$1 - \sum_{j=1}^{q} P[|\hat{S}_{nj}| > t] \le 1 - P(\hat{S}_n \notin \mathcal{S}_t) = P(\hat{S}_n \in \mathcal{S}_t).$$

Moreover,

$$P(\hat{S}_n \in \mathcal{S}_t) \ge 1 - \sum_{j=1}^{q} P\left(\frac{|V_j|}{\hat{\sigma}_j} > t\right) - \sum_{j=1}^{q}\left[P(|\hat{S}_{nj}| > t) - P\left(\frac{|V_j|}{\hat{\sigma}_j} > t\right)\right]$$

$$= 1 - \sum_{j=1}^{q} P\left(\frac{|V_j|}{\sigma_j} > \frac{\hat{\sigma}_j}{\sigma_j} t\right) - \sum_{j=1}^{q}\left[P\left(|S_{nj}| > \frac{\hat{\sigma}_j}{\sigma_j} t\right) - P\left(\frac{|V_j|}{\sigma_j} > \frac{\hat{\sigma}_j}{\sigma_j} t\right)\right].$$

But

$$\left[P\left(|S_{nj}| > \frac{\hat{\sigma}_j}{\sigma_j} t\right) - P\left(\frac{|V_j|}{\sigma_j} > \frac{\hat{\sigma}_j}{\sigma_j} t\right)\right] = \left[P\left(\frac{V_j}{\sigma_j} \le \frac{\hat{\sigma}_j}{\sigma_j} t\right) - P\left(S_{nj} \le \frac{\hat{\sigma}_j}{\sigma_j} t\right)\right]$$

$$- \left[P\left(\frac{V_j}{\sigma_j} < -\frac{\hat{\sigma}_j}{\sigma_j} t\right) - P\left(S_{nj} < -\frac{\hat{\sigma}_j}{\sigma_j} t\right)\right].$$

By the triangle inequality,

$$\left| P\left(|S_{nj}| > \frac{\hat{\sigma}_j}{\sigma_j} t \right) - P\left(\frac{|V_j|}{\sigma_j} > \frac{\hat{\sigma}_j}{\sigma_j} t \right) \right| = \left| P\left(\frac{V_j}{\sigma_j} \le \frac{\hat{\sigma}_j}{\sigma_j} t \right) - P\left(S_{nj} \le \frac{\hat{\sigma}_j}{\sigma_j} t \right) \right|$$

$$+ \left| P\left(\frac{V_j}{\sigma_j} < -\frac{\hat{\sigma}_j}{\sigma_j} t \right) - P\left(S_{nj} < -\frac{\hat{\sigma}_j}{\sigma_j} t \right) \right|.$$

By the Berry-Esséen inequality with the constant of Shevstova (2011),

$$\sup_t \left| P\left(|S_{nj}| > \frac{\hat{\sigma}_j}{\sigma_j} t \right) - P\left(\frac{|V_j|}{\sigma_j} > \frac{\hat{\sigma}_j}{\sigma_j} t \right) \right| \le 0.9496 \frac{m_{j3}}{n^{1/2}}.$$

Therefore,

$$P(\hat{S}_n \in \mathcal{S}_t) \ge 1 - \sum_{j=1}^{q} P(|V_j| > \hat{\sigma}_j t) - 0.9496 n^{-1/2} \sum_{j=1}^{q} m_{j3}$$

.

$$(9.14) \qquad \ge 1 - \sum_{j=1}^{d} P(|V_j| > \hat{\sigma}_j t) - 0.9496 \frac{q m_3}{n^{1/2}}.$$

Step 2: The upper bound.

$$P(\hat{S}_n \in \mathcal{S}_t) = P\left[\bigcap_{j=1}^{q} (|\hat{S}_{nj}| \le t) \right] \le \sum_{j=1}^{q} P(|\hat{S}_{nj}| \le t)$$

Therefore,

$$P(\hat{S}_n \in \mathcal{S}_t) \le \sum_{j=1}^{q} P(|V_j| \le \hat{\sigma}_j t) + \sum_{j=1}^{q} \left[P(|\hat{S}_{nj}| \le t) - P(|V_j| \le \hat{\sigma}_j t) \right].$$

By the Berry-Esséen inequality

$$(9.15) \quad P(\hat{S}_n \in \mathcal{S}_t) \le \sum_{j=1}^{q} P\left(\frac{|V_j|}{\hat{\sigma}_j} \le t \right) + 0.9496 \frac{q m_3}{n^{1/2}}$$

uniformly over t.

The lemma follows by combining (9.14) and (9.15). Q.E.D.

Theorem 9.1: For any $\tau > 0$, define

$$\tilde{\Xi}_{n\tau} = \min \begin{cases} q 2^{q+1} \tilde{r}(\tau)] \\ \frac{1}{\sqrt{2}} \{\tilde{r}(\tau) - \log[1 - \tilde{r}(\tau)]\}^{1/2} \end{cases}.$$

(i) Let assumptions 1-3 hold. Treat $\hat{\sigma}_j$ ($j = 1,...,q$) as non-stochastic. For any $\tau > 0$ such that $\max[\tilde{r}(\tau), r(\tau)] < 1$, then uniformly over t

$$1-\sum_{j=1}^{q} P\left(\frac{|\hat{V}_j|}{\hat{\sigma}_j} > t\right) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right) - q\tilde{\Xi}_{n\tau} - 0.9496\frac{qm_3}{n^{1/2}} \le P(\hat{S}_n \in \mathcal{S}_t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right)$$

$$\le \sum_{j=1}^{q} P(|\hat{V}_j| \le \hat{\sigma}_j t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right) + q\tilde{\Xi}_{n\tau} + 0.9496\frac{qm_3}{n^{1/2}}$$

with probability at least $1-4q^2 e^{-\tau}$.

(ii) Define

$$w_n = \inf_{\tau:\, \max[\tilde{r}(\tau), r(\tau)] < 1} (q\Xi_{n\tau} + 4q^2 e^{-\tau}).$$

Under the conditions of part (i) and uniformly over t,

$$1-\sum_{j=1}^{q} P\left(\frac{|\hat{V}_j|}{\hat{\sigma}_j} > t\right) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right) - w_n - 0.9496\frac{qm_3}{n^{1/2}} \le P(\hat{S}_n \in \mathcal{S}_t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right)$$

$$\le \sum_{j=1}^{q} P(|\hat{V}_j| \le \hat{\sigma}_j t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right) + w_n + 0.9496\frac{qm_3}{n^{1/2}}.$$

(iii) Let assumption 1 hold. Treat $\hat{\sigma}_j$ $(j = 1,...,q)$ as non-stochastic. If every component of $Z[Y - g(X,\theta_0)]$ is sub-Gaussian with variance proxy σ_{SG}^2, then uniformly over t,

$$(9.16)\quad 1-\sum_{j=1}^{q} P(|\hat{V}_j| > \hat{\sigma}_j t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right) - qw_n - 0.9496\frac{qm_3}{n^{1/2}}$$

$$\le P(\hat{S}_n \in \mathcal{S}_t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right)$$

$$\le \sum_{j=1}^{q} P(|\hat{V}_j| \le \hat{\sigma}_j t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right) + qw_n + 0.9496\frac{qm_3}{n^{1/2}}.$$

Proof: Part (i). By Lemma 9.4

$$(9.17) \quad 1-\sum_{j=1}^{q} P(|\hat{V}_j|>\hat{\sigma}_j t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right) + \sum_{j=1}^{q}[P(|\hat{V}_j|>\hat{\sigma}_j t) - P(|V_j|>\hat{\sigma}_j t)] - 0.9496\frac{qm_3}{n^{1/2}}$$

$$\le P(\hat{S}_n \in \mathcal{S}_t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right)$$

$$\le \sum_{j=1}^{q} P(|\hat{V}_j| \le \hat{\sigma}_j t) - P\left(\max_j \frac{|\hat{V}_j|}{\hat{\sigma}_j} \le t\right) - \sum_{j=1}^{q}[P(|\hat{V}_j|>\hat{\sigma}_j t) - P(|V_j|>\hat{\sigma}_j t)] + 0.9496\frac{qm_3}{n^{1/2}}.$$

By Lemma 9.3

$$(9.18) \quad |P(|\hat{V}_j|>\hat{\sigma}_j \tau) - P(|V_j|>\hat{\sigma}_j \tau)| \le \tilde{\Xi}_{n\tau}$$

with probability at least $1-4q^2 e^{-\tau}$ if $\max[\tilde{r}(\tau), r(\tau)] < 1$. Part (i) follows by substituting (9.18) into (9.17).

Part (ii) is a straightforward consequence of part (i).

Part (iii). Define κ_n as in step 3 of the proof of Corollary 4.3. As in the proof of Lemma 9.3,

$$\sup_{\tau} |P(\hat{V}_j \le \hat{\sigma}_j \tau) - P(V_j \le \hat{\sigma}_j \tau)| \le 4\frac{|\hat{\sigma}_j^2 - \sigma_j^2|}{\sigma_j^2}.$$

As in Lemma 9.3 and Corollary 4.2, $\hat{\sigma}_j$ is treated as non-stochastic in this inequality. As in step 3 of the proof of Corollary 4.3,

$$P\left[4\frac{|\hat{\sigma}_\xi^2 - \sigma_\xi^2|}{\sigma_\xi^2} > 16\sqrt{2}\sigma_{SG}^2\left(\frac{t}{n}\right)^{1/2}\right] \le 3e^{-t}.$$

if $t \le \kappa_n$. Thus, if $t \le \kappa_n$,

$$\sup_{\tau} |P(\hat{V}_j \le \hat{\sigma}_j \tau) - P(V_j \le \hat{\sigma}_j \tau)| \le \min_{t \le \kappa_n}\left[16\sqrt{2}\sigma_{SG}^2\left(\frac{t}{n}\right)^{1/2} + 3e^{t}\right] \equiv w_n.$$

Therefore, (9.16) follows from substituting this result into (9.17). Q.E.D.

TABLE 1: NUMERICAL EXAMPLES OF THE BOUND OF COROLLARY 4.2[a]

n	Distr. of U	m_3	ℓ	Bound	Bound with known σ_{FWL}^2
1000	$U[-2,2]$	2.073	2.828	0.641	0.062
1000	Laplace	2.394	1.414	0.346	0.072
10,000	$U[-2,2]$	2.073	2.828	0.321	0.020
10,000	Laplace	2.394	1.414	0.103	0.023
100,000	$U[-2,2]$	2.073	2.828	0.115	0.006
100,000	Laplace	2.394	1.414	0.035	0.007

a. Bounds are from Corollary 3.3 with $\beta = \gamma = 0$, $Z_X \sim N(0,1)$, and U distributed as $U[-2,2]$ or Laplace.

Table 2: Empirical Probabilities of Rejecting Correct Null Hypotheses at the Nominal 0.05 Level

Distr.	n	$q = 1$	$q = 2$	$q = 5$	$q = 10$
Uniform	100	0.053	0.051	0.050	0.039
	1000	0.056	0.050	0.050	0.051
Skewed	100	0.055	0.049	0.045	0.044
	1000	0.051	0.052	0.051	0.048
Bimodal	100	0.056	0.051	0.041	0.035
	1000	0.049	0.049	0.053	0.051
Laplace	100	0.052	0.048	0.039	0.027
	1000	0.050	0.050	0.048	0.050
$t(10)$	100	0.049	0.050	0.039	0.034
	1000	0.049	0.047	0.050	0.050
Diff. betw. Lognormals	100	0.048	0.037	0.023	0.012
	1000	0.052	0.045	0.045	0.034

Table 3: Powers of the T_n and Stock-Wright Tests of a Simple Null Hypothesis at the Nominal 0.05 Level

Distr.	n	β_0	c	$q=1$ T_n	$q=1$ SW	$q=2$ T_n	$q=2$ SW	$q=5$ T_n	$q=5$ SW	$q=10$ T_n	$q=10$ SW
Uniform	100	1.0	0.50	0.607	0.604	0.803	0.805	0.964	0.964	0.996	0.996
	1000	0.20	0.50	0.653	0.652	0.857	0.856	0.990	0.990	1.000	1.000
Skewed	100	1.0	0.50	0.429	0.429	0.593	0.593	0.831	0.835	0.947	0.955
	1000	0.20	0.50	0.459	0.457	0.648	0.649	0.908	0.909	0.991	0.992
Bimodal	100	1.0	0.50	0.272	0.271	0.355	0.358	0.551	0.562	0.712	0.753
	1000	0.20	0.50	0.276	0.274	0.385	0.385	0.636	0.638	0.852	0.852
Laplace	100	1.0	0.50	0.484	0.482	0.643	0.062	0.853	0.886	0.949	0.972
	1000	0.20	0.50	0.488	0.488	0.687	0.689	0.931	0.936	0.996	0.996
$t(10)$	100	1.0	0.50	0.646	0.645	0.827	0.837	0.971	0.974	0.996	0.999
	1000	0.20	0.50	0.687	0.685	0.875	0.875	0.994	0.995	1.000	1.000
Diff. betw. Lognormals	100	1.0	0.50	0.168	0.167	0.212	0.241	0.260	0.410	0.303	0.650
	1000	0.20	0.50	0.164	0.162	0.198	0.208	0.288	0.338	0.392	0.525

Table 4: Powers of the T_n and Stock-Wright Tests of a Simple Null Hypothesis at the Nominal 0.05 Level

Distr.	n	β_0	c	$q=1$ T_n	$q=1$ SW	$q=2$ T_n	$q=2$ SW	$q=5$ T_n	$q=5$ SW	$q=10$ T_n	$q=10$ SW
Uniform	100	1.0	0.25	0.196	0.194	0.275	0.266	0.425	0.423	0.569	0.597
	1000	0.20	0.25	0.223	0.221	0.305	0.300	0.498	0.495	0.714	0.714
Skewed	100	1.0	0.25	0.151	0.149	0.188	0.189	0.256	0.282	0.352	0.444
	1000	0.20	0.25	0.153	0.151	0.202	0.200	0.304	0.310	0.465	0.469
Bimodal	100	1.0	0.25	0.109	0.108	0.122	0.122	0.149	0.179	0.184	0.267
	1000	0.20	0.25	0.105	0.102	0.126	0.124	0.177	0.180	0.245	0.253
Laplace	100	1.0	0.25	0.166	0.162	0.200	0.215	0.278	0.341	0.368	0.525
	1000	0.20	0.25	0.162	0.159	0.214	0.214	0.325	0.338	0.495	0.521
$t(10)$	100	1.0	0.25	0.222	0.220	0.297	0.303	0.446	0.484	0.581	0.688
	1000	0.20	0.25	0.234	0.233	0.318	0.320	0.520	0.524	0.740	0.751
Diff. betw. Lognormals	100	1.0	0.25	0.078	0.076	0.080	0.093	0.072	0.132	0.058	0.232
	1000	0.20	0.25	0.082	0.082	0.829	0.871	0.092	0.104	0.099	0.137

Table 5: Powers of the T_n and Stock-Wright Tests of a Composite Null Hypothesis at the Nominal 0.05 Level

Distr.	n	β_1, β_2	c	$q=1$ T_n	$q=1$ SW	$q=2$ T_n	$q=2$ SW	$q=5$ T_n	$q=5$ SW	$q=10$ T_n	$q=10$ SW
Uniform	100	1.0	0.50	0.610	0.608	0.801	0.805	0.967	0.968	0.996	0.996
	1000	1.0	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	0.20	0.50	0.659	0.659	0.856	0.856	0.992	0.992	1.00	1.00
Skewed	100	1.0	0.50	0.423	0.421	0.587	0.592	0.828	0.839	0.944	0.954
	1000	1.0	0.50	0.999	0.999	1.00	1.00	1.00	1.00	1.00	1.00
	1000	0.20	0.50	0.460	0.462	0.643	0.642	0.912	0.912	0.993	0.994
Bimodal	100	1.0	0.50	0.263	0.260	0.364	0.361	0.548	0.566	0.713	0.756
	1000	1.0	0.50	0.987	0.988	1.00	1.00	1.00	1.00	1.00	1.00
	1000	0.20	0.50	0.278	0.276	0.385	0.382	0.628	0.629	0.854	0.856
Laplace	100	1.0	0.50	0.470	0.470	0.629	0.642	0.849	0.883	0.945	0.971
	1000	1.0	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	0.20	0.50	0.491	0.488	0.681	0.687	0.936	0.938	0.995	0.995
$t(10)$	100	1.0	0.50	0.651	0.648	0.822	0.824	0.971	0.974	0.995	0.997
	1000	1.0	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1000	0.20	0.50	0.681	0.679	0.881	0.882	0.994	0.994	1.00	1.00
Diff. betw. Lognormals	100	1.0	0.50	0.167	0.167	0.205	0.234	0.254	0.401	0.307	0.638
	1000	1.0	0.50	0.794	0.793	0.936	0.942	0.995	0.998	0.999	1.00
	1000	0.20	0.50	0.155	0.152	0.197	0.205	0.286	0.336	0.389	0.523

Table 6: Powers of the T_n and Stock-Wright Tests of a Composite Null Hypothesis at the Nominal 0.05 Level

Distr.	n	β_1, β_2	π	$q=1$ T_n	$q=1$ SW	$q=2$ T_n	$q=2$ SW	$q=5$ T_n	$q=5$ SW	$q=10$ T_n	$q=10$ SW
Uniform	100	1.0	0.25	0.207	0.205	0.283	0.281	0.420	0.427	0.572	0.593
	1000	1.0	0.25	0.955	0.955	0.998	0.998	1.00	1.00	1.00	1.00
	1000	0.20	0.25	0.220	0.217	0.301	0.299	0.497	0.494	0.708	0.703
Skewed	100	1.0	0.25	0.145	0.144	0.182	0.184	0.257	0.290	0.345	0.429
	1000	1.0	0.25	0.817	0.816	0.959	0.959	0.999	0.999	1.00	1.00
	1000	0.20	0.25	0.155	0.155	0.202	0.201	0.312	0.313	0.458	0.463
Bimodal	100	1.0	0.250	0.106	0.102	0.0.125	0.127	0.150	0.177	0.190	0.267
	1000	1.0	0.25	0.559	0.557	0.763	0.764	0.970	0.970	0.999	0.999
	1000	0.20	0.25	0.108	0.106	0.126	0.125	0.179	0.177	0.250	0.256
Laplace	100	1.0	0.25	0.164	0.160	0.202	0.211	0.282	0.352	0.362	0.523
	1000	1.0	0.25	0.843	0.843	0.972	0.972	1.00	1.00	1.00	1.00
	1000	0.20	0.25	0.162	0.162	0.218	0.215	0.334	0.345	0.490	0.516
$t(10)$	100	1.0	0.25	0.223	0.222	0.292	0.300	0.444	0.448	0.590	0.685
	1000	1.0	0.25	0.962	0.962	0.999	0.999	1.00	1.00	1.00	1.00
	1000	0.20	0.25	0.239	0.237	0.320	0.319	0.526	0.527	0.744	0.752
Diff. betw. Lognormals	100	1.0	0.25	0.074	0.070	0.079	0.092	0.072	0.134	0.062	0.229
	1000	1.0	0.25	0.300	0.299	0.406	0.420	0.641	0.698	0.830	0.913
	1000	0.20	0.25	0.076	0.076	0.081	0.086	0.090	0.105	0.096	0.139

REFERENCES

Acemoglu, D., S. Johnson, and J.A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *The American Economic Review*, 91, 1369-1401.

Anderson, T.W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20, 46-63.

Andrews, D.W.K. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80, 2153-2211.

Andrews, D.W.K. and P. Guggenberger (2018). Identification- and Singularity-Robust Inference for Moment Condition Models. Cowles Foundation discussion paper no. CFDP 1978R.

Andrews, D.W.K. and V. Marmer (2008). Exactly distribution-free inference in instrumental variables regression with possibly weak instruments. *Journal of Econometrics*, 142, 183-200.

Andrews, D.W.K., V. Marmer, and Z. Yu (2019). Optimal inference in the linear IV model. *Quantitative Economics*, 10, 457-485.

Andrews, D.W.K., M.J. Moreira, and J.H. Stock (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica*, 74, 715-752.

Andrews, D.W.K. and G. Soares (2007). Rank tests for instrumental variables regression with weak instruments. *Econometric Theory*, 23, 1033-1082.

Andrews, D.W.K. and J.H. Stock (2007a). Inference with weak instruments. In *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress*, Vol. 3, R. Blundell, W. Newey, and T. Persson, eds. Cambridge: Cambridge University Press, pp. 122-173.

Andrews, D.W.K. and J.H. Stock (2007b). Testing with many weak instruments. *Journal of Econometrics*, 138, 24-46.

Andrews, I. and A. Mikusheva (2016). Conditional inference with a functional nuisance parameter. *Econometrica*, 84, 1571-1612.

Angrist, J. and A. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106, 979-1014.

Bekker, P.A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62, 657-681.

Bentkus, V. (2003). On the dependence of the Berry-Esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113, 385-402.

Blundell, R., J.L. Horowitz, and M. Parey (2012). Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics*, 3, 29-51.

Bound, J., D.A. Jaeger, and R.M. Baker (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.

Brent, R.P., J.H. Osborn, and W.D. Smith (2015). Note on best possible bounds for determinants of matrices close to the identity matrix. *Linear Algebra and Its Applications*, 466, 21-26.

Bugni, F.A., I.A. Canay, and X. Shi (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, 8, 1-38.

Carrasco, M. and G. Tchuente (2016). Regularization based Anderson Rubin tests for many instruments. Working paper, Department of Economics, University of Montreal.

Cheng, X. (2015). Robust inference in nonlinear models with mixed identification strength. *Journal of Econometrics*, 189, 207-228.

Chernozhukov, V., D. Chetverikov, and K. Kato (2017). Central limit theorems and bootstrap in high dimensions, arXiv:1412.3661v4 [math.ST], *Annals of Probability*, forthcoming.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics*, 41, 2786-2819.

Chernozhukov, V., C. Hansen, and M. Jansson (2009). Finite sample inference for quantile regression models. *Journal of Econometrics*, 152, 93-103.

Dasgupta, A. (2008). *Asymptotic Theory of Statistics and Probability.* New York: Springer.

Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica*, 73, 1351-1365.

Guggenberger, Kleibergen, Mavroeidis (2019). A more powerful subvector Anderson Rubin test in linear instrumental variables regression; *Quantitative Economics*, 10, 487-526.

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012). On the asymptotic sizes of subset Anderson-Rubin and Lagrange multiplier tests in linear instrumental variables regression. *Econometrica*, 80, 2649-2666.

Guggenberger, P. and R.J. Smith (2005). Generalized empirical likelihood estimators and tests under partial, weak, and strong identification. *Econometric Theory*, 21, 667-709.

Hansen, C., J. Hausman, and W. Newey (2008). Estimation with many instrumental variables. *Journal of Business and Economic Statistics*, 26, 398-422.

Holly, A. and P.C.B. Phillips (1979). A saddlepoint approximation to the distribution of the *k*-class estimator of a coefficient in a simultaneous system. *Econometrica*, 47, 1527-1547.

Horowitz, J.L. (2006). Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica*, 74, 521-538.

Horowitz, J.L. and S. Lee (2009). Testing a parametric quantile-regression model with an endogenous explanatory variable against a nonparametric alternative. *Journal of Econometrics*, 142, 141-152.

Imbens, G.W., R.H. Spady, and P. Johnson (1998). Information theoretic approaches to inference in moment condition models. *Econometrica*, 66, 333-357.

Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, 40, 633-643.

Jun, S.J. (2008). Weal identification robust tests in an instrumental quantile model. *Journal of Econometrics*, *Journal of Econometrics*, 144, 118-138.

Kitamura, Y. and M. Stutzer (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65, 861-874.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70, 1781-1803.

Moreira, M.J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, 71, 1027-1048.

Moreira, M.J. (2009). Tests with correct size when instruments can be arbitrarily weak. *Journal of Econometrics*, 152, 131-140.

Nelson, C.R. and R. Startz (1990a). The distribution of the instrumental variable estimator and its t ratio when the instrument is a poor one. *Journal of Business*, 63, S124-S140.

Nelson, C.R. and R. Startz (1990b). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica*, 58, 967-976.

Newey, W.K. and R.J. Smith (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72, 219-255.

Newey, W.K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, 77, 687-719.

Phillips, P.C.B. (1983). Exact small sample theory in the simultaneous equations model. In *Handbook of Econometrics*,Vol. 1, Z. Griliches and M.D. Intrilligator, eds., North-Holland Publishing Company, pp. 449-516.

Raič, M. (2019). A multivariate Berry-Essen theorem with explicit constants. *Bernoulli*, 25, 2824-2853.

Rothenberg, T.J. (1984). Approximating the distributions of econometric estimators and test statistics. In *Handbook of Econometrics*,Vol. 2, Z. Griliches and M.D. Intrilligator, eds., North-Holland Publishing Company, pp. 881-935.

Shevstova, I. (2011). On the absolute constants in the Berry Esseen type inequalities for identically distributed summands. arXiv:1111.6554

Spokoiny, V. (2012). Parametric estimation. Finite sample theory. *Annals of Statistics*, 40, 2877-2909.

Spokoiny, V. and M. Zhilova (2015). Bootstrap confidence sets under model misspecification. *Annals of Statistics*, 43, 2653-2675.

Staiger, D. and J.H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65, 557-586.

Stock. J.H. and J.H. Wright (2000). GMM with weak identification. *Econometrica*, 68, 1055-1096.

Wang, J. and E. Zivot (1998). Inference on structural parameters in instrumental variables regression with weak instruments. *Econometrica*, 66, 1389-1404.